ON WEAK AND STRONG CONVERGENCE RATE FOR THE HESTON STOCHASTIC VOLATILITY MODEL

by

Chao Zheng

Submitted for the degree of

Doctor of Philosophy

Department of Mathematics

School of Mathematical and Computer Sciences

Heriot-Watt University

June 2016

Abstract

The Heston stochastic volatility model is one of the most fundamental models in mathematical finance. Recently, many numerical schemes have been developed for the Heston model. However, in the literature, there is no weak or strong convergence rate obtained for the full parameter regime. In this PhD thesis, we shall focus on the numerical scheme that simulates the variance process exactly and applies the stochastic trapezoidal rule to approximate the time integral of the variance process in the SDE of the logarithmic asset process. Our goal is to obtain the weak and strong convergence rates of such a numerical scheme for the Heston model.

The weak convergence rate is of traditional interest, because it is an important measure on how fast the bias of a numerical scheme decays. We prove that the numerical scheme we consider converges at rate two for the whole parameter regime, and the test function can be any polynomial of the logarithmic asset process. The rate is consistent with the standard rate of the stochastic trapezoidal rule, although the Lipschitz assumption is not satisfied. The strong convergence analysis is meaningful in the framework of Multi-level Monte Carlo (MLMC). The MLMC can be regarded as a variance reduction technique for numerical schemes on SDEs, as long as there is a MLMC estimator with a good strong convergence rate. We establish efficient MLMC estimators, separately for the path-independent and path-dependent simulations. We are able to provide the strong convergence rates in both situations.

Acknowledgements

Before ending up my research as a PhD student at Heriot-Watt University, I would like to take this opportunity to say thank you to these who have helped me intensively during this unforgettable journey. It is them that make my PhD life fruitful and enjoyable. My first gratitude goes to my first supervisor, Dr Anke Wiese, and my second supervisor, Dr Simon Malham, who have guided me to do research. I appreciate useful comments from them that have improved the presentation of the thesis. In addition, I have learnt a lot from Prof Mike Giles through discussions on the Multilevel Monte Carlo part of the thesis, with many thanks. I have benefited a lot from my friends Iain Ritchie and Sergio Inglima, who have spent much time to proofread my thesis. Except for that, I have learnt much from Iain about industrial applications of quantitative finance, and from Sergio on mathematics in general. I appreciate the assistance from Max Lau very much, who has kindly shared his wonderful research experience with me. Furthermore, I am quite enjoying the atmosphere in my office, full of pleasure and fun, which is due to my officemates, Iain, Junye, Houry, Neil, and Syahilla. We have worked together for three years since the beginning, although on different research topics. Finally, I owe a lot to my parents, who have provided me the best they could, and I owe them for their great support and encouragement.

Contents

List of Figures

Chapter 1

Introduction

Over the last few decades, mathematical finance has been becoming a rapidly developing field, as the derivatives have becoming increasing important in finance. 'The value of the assets underlying outstanding derivatives transactions is several times the world gross domestic product (Hull 2012, Chapter 1).' Many models have thus been established to fit the real data in various financial markets. Among them, of particular theoretical and practical interest is the Heston stochastic volatility model (Heston 1993), which is usually called the Heston model for convenience. This model has huge applications in the markets of equity, fixed income, and foreign exchange, and it is one of the fundamental models in mathematical finance. Numerous models are developed based on the Heston model, for example, the SVJ model proposed by Bates (1996). This model extends the Heston model by adding a jump diffusion.

The Heston model uses a two-dimensional stochastic differential equation to describe the evolution of two stochastic processes, the asset process and the variance process. The variance process at any time follows a non-central chi-squared distribution, and it is sometimes called the CIR process. The asset process can be written as the exponential of some integrals of the variance process. Under the Heston model, the price of the standard European option has a semi-closed formula, that is easy to calculate. However, the prices of the majority of options cannot be written in closed or semi-closed forms. Therefore, Monte Carlo techniques become essential. Furthermore, since there is no analytical solution of the Heston model, one has to approximate it by some numerical schemes. However, the traditional numerical methods, as those in Kloeden & Platen (1999) or Milstein (1995), are either ill-defined or inefficient for the Heston model. This phenomenon has been noted in many articles, for example, in Glasserman & Kim (2011). For this reason, in the recent years, many methods have been published for the Monte Carlo simulation of the Heston model.

Nevertheless, none of these numerical schemes has an analytical weak convergence rate in the literature. The weak convergence rate indicates how fast the bias of a numerical scheme decays with the step size, and thus it is an important measure on the simulation efficiency. In the literature, the weak convergence rate for the Heston model can only be assessed numerically, which has no theoretical guarantee. The first challenge we face is what is the analytical weak convergence rate of a numerical scheme for the Heston model. As the global Lipschitz condition is not satisfied in the Heston SDE, the traditional analysis as in Kloeden & Platen (1999) is not applicable. Actually, the weak convergence rate for the Heston model is a long-standing problem, as discussed in a recent review by Kloeden & Neuenkirch (2012), although there are plenty of results of non-Lipschitz analysis for general SDEs. Even for a simple Euler scheme for the Heston model, the weak convergence rate is missing in the literature.

The second challenge is how to combine the Multi-level Monte Carlo (MLMC), introduced by Giles (2008*b*), with a numerical scheme for the Heston model. This may require the construction of a non-standard MLMC estimator. The MLMC can be regarded as a variance reduction technique for numerical methods on SDEs. The computational complexity of the MLMC depends highly on the convergence rate of the variance of underlying MLMC estimator. There are several MLMC applications in the literature for the Heston model, for example Giles (2008*b*). However, they are all restricted to the parameter regime, where the zero boundary of the variance process is not attainable. As discussed in Andersen (2008), it is often observed that in the Heston model, the zero boundary is attainable and reflecting.

In this thesis, we shall focus on the numerical scheme for the Heston model that simulates the variance process exactly, and applies the stochastic trapezoidal rule to approximate the time integral of the variance process within the SDE of the logarithmic asset process. This is consistent with several efficient schemes in the existing literature. Since the variance process at any time follows a non-central chi-squared distribution, some authors considered either to approximate it with some random variables that can be easily simulated, or to simulate it almost exactly. In the literature, we have Andersen (2008), Van Haastrecht & Pelsser (2010) and Malham & Wiese (2013). All of these methods use the stochastic trapezoidal discretization for the SDE of the logarithmic asset process, and they are demonstrated to be highly efficient for the Heston model in the numerical tests with realistic model parameters.

For the first challenge, we prove the weak convergence rate of the stochastic trapezoidal rule is two, for the full parameter regime, provided the variance process is exactly simulated. This rate is consistent with the standard rate of the stochastic trapezoidal rule. The error criteria we use can be the difference of any polynomial function of the logarithmic asset process and that of its approximation. In order to prove this, we develop the theory of the stochastic trapezoidal rule in a more general context. We impose conditions on the expectation function of the product of some stochastic processes at different time, which is by nature a deterministic function. The major assumption is that the expectation function is a twice continuously differentiable function of time, which is easy to satisfy for many stochastic processes. In the literature, the usual analysis of SDEs is based on the analysis of model coefficients. The traditional analysis requires the Lipschitz condition and the linear growth condition, which are not satisfied for many SDEs. However, our analysis focuses on the structure of the expectation function and there is no direct assumptions on the coefficient. Furthermore, we transfer the error analysis to that of a trapezoidal rule on a multiple integral, which is deterministic. This implies that many powerful tools in analysing a more general quadrature rule on a deterministic multiple integral can be potentially applied or extended to solve problems arising in SDEs. Finally, the result is extended to the same numerical scheme for the SVJ model, also with the same weak convergence rate under the full parameter regime.

For the second challenge, we aim to implement efficient MLMC estimators that apply with no restriction on the parameters. We separate the simulations for path-independent options and for path-dependent options, because the underlying numerical schemes to price these two types of options differ slightly. For a path-independent option, the price of such an option only relies on the asset process at maturity. For a path-dependent option, the price depends on the path of the asset process before maturity. We shall call the simulation for path-independent options the path-independent simulation and that for path-dependent options path-dependent simulation. The MLMC estimators are constructed in both situations. These MLMC estimators are all bias-free, and are applicable without any further restrictions on the payoff function of an option. Since the computational complexity of the MLMC depends on the convergence rate of the variance of the MLMC estimator, it is important to evaluate the convergence rate. In the theoretical part, the analysis is based on the logarithmic asset price, which is typically for put options of Lipschitz payoff functions. We show that, in the path-independent simulation, the convergence rate is two for the full parameter regime. In the path-dependent simulation, the rate is one under some constraint on the parameters and it is half in all parameter regimes. The proof again concentrates on the structure of the variance process. It is different from the usual approach in the literature, which focuses on coefficients of SDEs. In the numerical part, we perform tests for the standard European call option, based on several sets of realistic parameters. In all the parameter regimes, the convergence rate is two in the path-independent simulation, and it is one in the path-dependent simulation. Finally, we extend these MLMC estimators for several Heston-like models, such as the SVJ model, the Heston model with stochastic interest rate, the Heston model with piecewise constant parameters, and the Heston model with CEV process. Here, in the Heston model with stochastic interest rate, we assume that the interest rate and variance process are mutually independent.

More generally, for the numerical solution of a high-dimensional SDE that has one or more components, the transition densities of which are known, it might be more convenient to simulate these components exactly or almost exactly and time-discretize the rest of the SDE. The approximation accuracy may depend on two factors. One is the properties of the stochastic process that can be exactly simulated. In this thesis, the major assumption we use is that the expectation of such a stochastic process is sufficiently smooth on a closed domain, and we are able to derive the weak and strong convergence rates [1] based on this property. The other is how we discretize the rest of the SDE. There are a number of efficient numerical methods falling into this category, but the mathematical understanding seems far from sufficient. We believe that there are plenty of space for future research, which would be helpful to understand and to develop more efficient numerical schemes for more general SDEs.

The thesis is organized as follows. In Chapter 2, we review the literature on the numerical methods for the Heston solution. Chapter 3 focuses on the stochastic trapezoidal rule for the Heston model, and discusses the weak convergence rate. Chapter 4 aims at constructing MLMC estimators and deriving the convergence rates of the variances. We conclude in Chapter 5, with some extensions and suggestions for the future research.

[1] We shall notice here that there is a clear difference between the definition of the strong convergence rate and that of the convergence rate of the variance of MLMC estimator. Usually, for a numerical scheme, the strong convergence rate is half of the rate of the variance, but this depends on which MLMC estimator we use. The definitions of both convergence rates are available in the next chapter.

Chapter 2

Monte Carlo Simulation for the Heston Model

This chapter has three sections. We start with section 2.1 discussing the Heston model. Section 2.2 is mainly about the fundamental theory of the stochastic differential equation, which provides an essential mathematical background for discussions in Section 2.3. Section 2.3 reviews the literature on numerical simulation methods for the Heston solution. This has become an active research area in mathematical finance, due to its huge applications for derivative pricing and risk management. Our focus in this thesis is always on the theoretical development, but we will cover several important industrial applications, that might be of interest to the reader.

2.1 Heston Stochastic Volatility Model

The history of mathematical finance dates back to Bachelier (1900), which is the first paper to discuss the use of Brownian motion to evaluate the stock price. However, there was a lack of mathematical rigour, as the theory of Brownian motion was not well-established at that time. Much effort, over many years, has been applied to establish the rigorous theory of the Brownian motion that we recognise today. This includes the celebrated work by Einstein (1905). It was not until 1973 that the fundamental theory of option pricing was constructed by Fischer Black and Myron Scholes. In their ground-breaking paper entitled 'The Pricing of Options and

Corporate Liabilities', the authors formulated the fair price of a standard European option, in the framework of arbitrage-free market. However, in the Black-Scholes model, the volatility for the underlying asset is assumed to be a constant, which is inconsistent with many empirical observations. For instance, in the foreign exchange market and in the equity market, the implied volatility[1] is usually a curve in the shape of 'smile' (Hull 2012, Section 19.2, 19.3).

For this reason, many stochastic volatility models were developed to better fit the implied volatility surface spotted in the market. The Heston model falls into this category, and it is described as 'Probably the most popular stochastic volatility model used in finance' (Platen & Bruti-Liberati 2010, Section 2.5). As discussed, for instance, in Andersen (2008), the Heston model is widely employed for the equity derivative pricing, fixed income derivative pricing and foreign exchange derivative pricing. The dynamics of the Heston model (Heston 1993) under the risk-neutral measure is as follow

$$\begin{aligned} \mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) \\ \mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1, \end{aligned}$$

in which $(S_t)_{t\geq 0}$ represents an asset price (a stock, an FX rate and so on), and $(V_t)_{t\geq 0}$ is the evolution of the variance. The constant r stands for the risk-neutral interest rate. The parameters k, θ, and σ are strictly positive constants, representing the speed of mean reversion, the long-run mean volatility, and the volatility of volatility respectively. Here, $(W_t^1)_{t\geq 0}$ and $(W_t^2)_{t\geq 0}$ are two independent Brownian motions, and ρ is the correlation taking value in $[-1, 1]$.

One of the most attractive research areas concerning the Heston model is to figure out the structure of its implied volatility surface, and to explore some of its asymptotic properties. The reader can visit Jacquier & Martini (2011) for a review. In this thesis, we shall, however, focus on the numerical simulation of the Heston

[1] The implied volatility usually refers to the value of the volatility within the Black-Scholes formula for the theoretical option price that is equal to the market price of a standard European option. It is standard to represent the market price of a standard European option by the implied volatility, despite that the Black-Scholes model may fail to work. This is because an implied volatility is much more informative than a single option price itself.

model, which can also be challenging. It is well-known that there is an unique strong solution of the Heston model, although the global Lipschitz condition is violated. However, the exact form of the solution is not available in the literature, and thus in many circumstances, one has to employ a time-discrete scheme for the numerical solution.

In financial applications, we consider an European-style option [2] with the underlying asset process S and the maturity T , the price of which is usually of the form

$$\mathbb{E}(D(0,T)f(S,T)|\mathcal{F}_0),$$

where $\mathbb{E}$ denotes the expectation under the risk-neutral measure, and $f : \mathbb{R} \times \mathbb{R}^+ \to \mathbb{R}^+$ is usually called the payoff function, which depends on the type of the option. Here, $D(0,T)$ is the discount factor, and generally we have $D(0,T) = e^{-\int_0^T r_t \mathrm{d}t}$, where $(r_t)_{t\geq 0}$ is the interest rate process. In the Heston model, the interest rate is constant, so the discount factor is also constant. Further, $\mathcal{F}_0$ is the initial condition. The following are some examples of options which are actively traded in the market. We let K be the strike price fixed, and S_t be the asset process S at time t, and we have

- Standard European call option (plain vanilla call option)

$$f(S,T) = (S_T - K)^+,$$

- Fixed strike Asian call option

$$f(S,T) = \left(\frac{1}{T}\int_0^T S_t \mathrm{d}t - K\right)^+,$$

- Floating strike Lookback call option

$$f(S,T) = \left(S_T - \inf_{0\leq t\leq T} S_t\right)^+,$$

[2]An European style option is the one that can only be exercised on expiration, whereas an American style option can be exercised before expiry. The computation of the value of an American option entails finding the optimal exercise rule, which is in general much more complicated (Glasserman 2003, Chapter 8).

- Up-and-out Barrier option

$$f(S,T) = (S_T - K)^+ 1_{\{\sup_{0\le t\le T} S_t \le B\}},$$

where B is the barrier level and $B > K$,

- Cash-or-nothing Digital call option

$$f(S,T) = 1_{\{S_T > K\}}.$$

The standard European option is a standard contract that gives option holders the right but not the obligation to buy or sell assets at a specified price on a specified date. It is one of the most common derivatives traded in the exchange. The other four options above are called exotic options, and they have more complicated payoff functions. These options are usually traded in the over-the-counter market.

With respect to the Heston model, the price of a standard European option, either a call or a put, can be expressed as a complex integral, that can be calculated efficiently, for instance, Carr & Madan (1999), Kahl & Jäckel (2005). These techniques provide a fast calibration of the Heston parameters for the market implied volatility. However, the majority of options cannot be priced in closed form, and Monte Carlo simulation becomes essential. There are two nice properties of the Heston model that are important for our simulation discussion.

One interesting property of the Heston model is that the variance process $(V_t)_{t\ge 0}$ can be simulated separately from the system of the Heston SDEs. In other words, one can first simulate $(V_t)_{t\ge 0}$, and then generate the sample path for the asset process $(S_t)_{t\ge 0}$. The cumulative distribution function of the variance process is known exactly. Specifically, V_t at time t follows a non-central Chi-square distribution, up to a scale factor, given V_u for some $u < t$. More precisely, we can write

$$V_t \overset{\mathrm{d}}{=} \frac{\sigma^2(1-e^{-k(t-u)})}{4k} \chi_d^2 \left(\frac{4ke^{-k(t-u)}}{\sigma^2(1-e^{-k(t-u)})} V_u \right),$$

where $\overset{\mathrm{d}}{=}$ means equality in distribution, $\chi_d^2(\lambda)$ denotes a non-central chi-square random variable with d degrees of freedom and non-centrality parameter λ, and $d = \frac{4\theta k}{\sigma^2}$. We notice that before Heston (1993), many properties of the variance process had already been known explicitly. Feller (1951) firstly introduced this stochastic process, and provided its density function. Cox, Ingersoll & Ross (1985) applied it to model the short-term interest rate, and the authors also derived the analytical formula for the bond price. For this reason, in the financial community, the variance process is now well-known as the CIR process. More properties about the CIR process and its extensions are available in Brigo & Mercurio (2006). By the Feller's classification of boundaries (Karlin & Taylor 1981, Section 15.6), when $2k\theta \geq \sigma^2$, the boundary zero of $(V_t)_{t\geq 0}$ is unattainable, and when $2k\theta < \sigma^2$, the origin is attainable and reflecting. Alfonsi (2010) claimed that the latter situation seldom occurs when the CIR process is used to represent the short interest rate, but they are often observed when it stands for the default intensity in credit risk or the stock volatility in the Heston model. Andersen (2008) provided several numerical tests for the Heston model based on realistic and challenging data from a variety of financial markets. They all fall into the latter category where the zero boundary is attainable and reflecting. However, as Glasserman & Kim (2011) pointed out, the standard Euler or Milstein methods, those in Kloeden & Platen (1999), often produce erratic results when applied to this model, particularly in the situation where the the zero boundary can be hit. Intuitively, this is because the square root in the Heston SDE becomes very sensitive when $(V_t)_{t\geq 0}$ stays in the neighbourhood of zero, which can make the truncation error from a numerical scheme difficult to control.

There is another notable property of the Heston model. Let $X_t = \ln(e^{-rt}S_t)$, and it follows that

$$X_{t+h} = X_t + \frac{\rho}{\sigma}(V_{t+h} - V_t - k\theta h) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_t^{t+h} V_s \mathrm{d}s + \sqrt{1-\rho^2}\sqrt{\int_t^{t+h} V_s \mathrm{d}s}\, N$$

for any positive h, where N is a standard Normal random variable, independent of the variance process. This formula is due to Broadie & Kaya (2006). Conditioned on the variance process simulated, one can use this formula for the numerical solution of $(S_t)_{t\geq 0}$ by taking a quadrature rule to approximate the time integral $\int_t^{t+h} V_s \mathrm{d}s$, for instance, the trapezoidal rule. Indeed, several numerical schemes for the Heston model in recent papers are based on the equation above, which we shall discuss later. Despite of its popularity in practice, there is very little analysis from a theoretical perspective. We shall show in this thesis that there are much more we can obtain from this equation, although its structure seems quite simple.

2.2 Theory of Stochastic Differential Equations

Before starting our review on simulation methods for the Heston model in the literature, we shall have some discussion on the theory of stochastic differential equations. It includes the definitions of the strong solution and the weak solution of a stochastic differential equation in general, and the criterion for the strong approximation and the weak approximation respectively. The criterion is essential when we talk about the rate of convergence in either a strong or weak sense.

Let us consider a d-dimensional stochastic differential equation

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t \tag{2.1}$$

written component-wise as

$$\mathrm{d}X_t^i = b_i(t, X_t)\mathrm{d}t + \sum_{j=1}^{l} \sigma_{ij}(t, X_t)\mathrm{d}W_t^j,$$

where $W = (W_t^1, W_t^2, ..., W_t^l)_{t\geq 0}$ is a l-dimensional Brownian motion and $b(t,x)$: $\mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma(t,x) : \mathbb{R}^+ \times \mathbb{R}^d \to \mathbb{R}^{d\times l}$, are Borel-measurable functions.

Definition 2.2.1. *A strong solution of the stochastic differential equation (2.1), on the given probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and with respect to the fixed Brownian motion W and the initial value ζ, is a process $(X_t)_{t\geq 0}$ with continuous sample paths and with the following properties:*

1. *$(X_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, where $\mathcal{F}_t = \sigma(\mathcal{G}_t \cup \mathcal{N})$ with $\mathcal{G}_t = \sigma(\zeta, W_s; 0 \leq s \leq t)$ and $\mathcal{N}$ the collection of $\mathbb{P}$-null sets;*

2. *$\mathbb{P}[X_0 = \zeta] = 1$;*

3. $\mathbb{P}[\int_0^t \{|b_i(s, X_s)| + \sigma_{ij}^2(s, X_s)\} \mathrm{d}s < \infty] = 1$ *holds for every* $i = 1, 2, ..., d$, $j = 1, 2, ..., l$, *and* $t \geq 0$;

4. *It holds almost surely that*

$$X_t = X_0 + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}W_s,$$

for all $t \geq 0$.

We take this definition from Karatzas & Shreve (1991), Section 5.2 and 5.3, as well as the weak solution below.

Definition 2.2.2. *A weak solution of the stochastic differential equation (2.1) is a triple* (X, W), $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_t)_{t \geq 0}$, *where*

1. $(\Omega, \mathcal{F}, \mathbb{P})$ *is a probability space, and* $(\mathcal{F}_t)_{t \geq 0}$ *is a filtration of sub-*σ*-fields of* $\mathcal{F}$ *satisfying the usual conditions;*

2. $X = \{X_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *is a continuous, adapted* $\mathbb{R}^d$*-valued process,* $W = \{W_t, \mathcal{F}_t; 0 \leq t < \infty\}$ *is an* l*-dimensional Brownian motion, and (3), (4) of Definition 2.2.1 are satisfied.*

A strong solution is by definition a weak solution, while in general, the reverse is not true. For example, we have the one-dimensional SDE, $\mathrm{d}X_t = sgn(X_t)\mathrm{d}W_t$, where $sgn(x) = 1, x > 0; sgn(x) = -1, x \leq 0$. This SDE has a weak solution, but it does not admit a strong solution (Karatzas & Shreve 1991, Section 5.3). Generally, the conditions of weak solution are easier to satisfy, as the filtration $(\mathcal{F}_t)_{t \geq 0}$ is not necessarily the augmentation of the filtration generated by the driving Brownian motion and by the initial condition $X_0 = \zeta$.

It is important that we should make sure the solution exists before a numerical scheme is considered to approximate it. For the variance process, there is an unique strong solution due to the Yamada-Watanabe theorem (Karatzas & Shreve 1991, Section 5.2.C), and the Heston solution can be written as the exponential of the integral of the variance process. Therefore, the Heston model admits an unique strong solution.

For numerical approximation of (2.1), if there is no analytical solution, we have to approximate it either in a strong sense or in a weak sense, depending on the specific application. Let $\hat{X}_{nh,h}$, $n = 0, 1, ..., T/h$, be a numerical approximation of the stochastic process $(X_t)_{t\geq 0}$ at time $t = nh$ through a time-discrete scheme with the time step size h. The definition of $\hat{X}_{nh,h}$ relies on the specific time-discrete scheme that is applied. For example, if we apply the Euler discretization on (2.1), the approximation of the solution has the form $\hat{X}_{(n+1)h,h} = \hat{X}_{nh,h} + b(nh, \hat{X}_{nh,h})h + \sigma(nh, \hat{X}_{nh,h})(W_{(n+1)h} - W_{nh})$ for all $n = 0, 1, ..., T/h-1$ with the initial value $\hat{X}_{0,h} = X_0$. The strong approximation usually refers to the approximation of the sample path $\{\hat{X}_{0,h}, \hat{X}_{h,h}, ..., \hat{X}_{T,h}\}$, to be close to the path-wise solution X with a given path of the Brownian motion W. To assess the efficiency of a time-discrete scheme of strong approximation, it is standard to consider $\mathbb{E}|\hat{X}_{T,h} - X_T|$, the convergence in $\mathbb{L}^1$ space, and there is a standard definition of the strong convergence rate γ below (Kloeden & Platen 1999).

Definition 2.2.3. *The numerical solution $\hat{X}_{T,h}$ converges in the strong sense with order $\gamma \in (0, \infty]$ if there exists a finite constant K and a positive constant L such that*

$$\mathbb{E}|\hat{X}_{T,h} - X_T| \leq Kh^{\gamma}$$

for any step size $h \in (0, L)$.

The traditional applications of the strong approximation include the direct simulation of trajectories of stochastic dynamical systems, the testing of parametric estimators and Markov chain filters (Kloeden & Platen 1999, Chapter 17). Recently, Giles (2008*b*) introduced the Multi-level Monte Carlo simulation, that also requires a strong approximation with a good convergence rate. However, for the MLMC, we shall use a different criterion for the convergence, because this criterion is more directly relevant to the error prediction as in theory of the MLMC. Let $\hat{X}_{T,Mh}$ be the numerical solution of X_T at time T with the time step size Mh, with M some positive integer. If the variance satisfies

$$\mathrm{Var}(\hat{X}_{T,h} - \hat{X}_{T,Mh}) = O(h^{\alpha}), \tag{2.2}$$

where α is independent of M, and then we call α the convergence rate of the variance of the MLMC estimator, or the convergence rate of the variance. The MLMC requires to construct a tie between the approximated solution $\hat{X}_{T,h}$ and $\hat{X}_{T,Mh}$, to make α as large as possible. We will discuss the MLMC specifically in Chapter 4. There is a strong relationship between the strong convergence rate γ and convergence rate of the variance α. Usually, for a numerical solution approximated by a time-discrete scheme that converges in a strong sense with the rate γ, we have $\alpha = 2\gamma$, but it also depends on how we construct the MLMC estimator based on this scheme. To get (2.2), we require

$$\mathbb{E}\left[(\hat{X}_{T,h} - \hat{X}_{T,Mh})^2\right] = O(h^\alpha).$$

Since

$$\mathbb{E}\left[(\hat{X}_{T,h} - \hat{X}_{T,Mh})^2\right] \leq 2\mathbb{E}\left[(\hat{X}_{T,h} - X_T)^2\right] + 2\mathbb{E}\left[(\hat{X}_{T,Mh} - X_T)^2\right],$$

a sufficient but not necessary condition for (2.2) is that

$$\mathbb{E}\left[(\hat{X}_{T,h} - X_T)^2\right] = O(h^\alpha).$$

The application of the Cauchy-Schwarz inequality implies

$$\mathbb{E}|\hat{X}_{T,h} - X_T| \leq \mathbb{E}^{1/2}\left[(\hat{X}_{T,h} - X_T)^2\right] = O(h^{\alpha/2}).$$

This explains why generally we have $\alpha = 2\gamma$.

Nevertheless, in many situations, it suffices to have a good approximation on the probability distribution of the solution of SDEs at a given final time, which is usually called the weak approximation. The analysis of the weak approximation relies on the test function we take. As discussed in Kloeden & Platen (1999), Page XXV, we can choose polynomials as test functions in the weak convergence criterion, although the class of test functions can be generalized slightly, to the class of continuously differentiable functions, with partial derivatives of polynomial growth. Here comes the standard criterion for the weak convergence, and the weak convergence rate β.

Definition 2.2.4. *The numerical solution $\hat{X}_{T,h}$ converges in the weak sense with order $\beta \in (0, \infty]$ if for any polynomial $p(.)$ there exists a finite constant K_p and a positive constant L_p such that*

$$|\mathbb{E}(p(\hat{X}_{T,h})) - \mathbb{E}(p(X_T))| \leq K_p h^\beta$$

for any step size $h \in (0, L_p)$.

From a more theoretical perspective, or in the review of functional analysis, the definition of the weak convergence can be generalized in the sense of weak topology in a Banach space, although it seems not directly relevant to practical applications so far.[3] In derivative pricing and sensitive analysis, one of the major goals is to approximate the option price, which is mathematically an expectation of a stochastic process at the maturity. Thus, it is usual in the literature to evaluate the pricing error by investigating the mean square error, the analysis of which is more relevant to that of the weak approximation. Specifically, suppose e is the true option price, and $\hat{e}$ is its approximation, we have the mean square error

$$\begin{aligned} \mathrm{MSE}(\hat{e}) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{e}_i - e\right)^2 \\ &= \frac{1}{n}\mathrm{Var}(\hat{e}) + \mathbb{E}^2(\hat{e} - e), \end{aligned}$$

where n is the sample size, and Var is for the variance. Under the condition that $\hat{e}$ converges to e in a strong sense, $\mathrm{Var}(\hat{e})$ converges to a constant $\mathrm{Var}(e)$, as the time step h goes to zero. Thus, the whole term $\frac{1}{n}\mathrm{Var}(\hat{e})$ should be $O(n^{-1})$. The analysis of $\mathbb{E}^2(\hat{e} - e)$ is exactly the analysis of weak convergence. Therefore, we have

$$\mathrm{MSE}(\hat{e}) = O(n^{-1}) + O(h^{2\beta}),$$

[3] If we consider the process $(X_t)_{t\in[0,T]}$ as a function on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we take $(\hat{X_t})_{t=h,2h,\ldots,T}$ as its approximation, and if we can define a Banach space B containing all these functions, the weak convergence means that for every continuous linear functional in the dual space B^* of B, denoted by $\varphi \in B^*$, such that the sequence $\varphi((\hat{X_t})_{t=h,2h,\ldots,T})$ converges to $\varphi((X_t)_{t\in[0,T]})$ as h approaches zero. The reader can go to Conway (2010), Section V.4, for more discussion on the weak convergence in a Banach space.

where β is the weak convergence rate. However, the rate of strong convergence can provide a lower bound for that of weak convergence, as the former is generally no larger than the latter.

We prefer a higher order numerical scheme, if not to consider other factors that may influence the simulation efficiency, for instance, whether the underlying numerical method is easy to implement. The simplest time-discrete scheme is the Euler scheme, with strong order 1/2, and weak order 1, under the usual conditions. The Milstein scheme, under the usual assumptions, improves the Euler scheme by raising the strong order to 1, while there is no improvement on the weak order. Higher order can be obtained by applying the Stochastic Taylor scheme, although the derivatives of various order of the drift and diffusion coefficients must be determined. To avoid this problem, one can approximate the derivatives by some finite difference methods, and get explicit or implicit schemes. However, the traditional convergence analysis of these schemes to guarantee the desired order are based on the usual conditions, such as the global Lipschitz and linear growth conditions on the coefficients, which are not satisfied for the Heston model. We shall say a function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies the global Lipschitz condition, if there exists a constant c, such that

$$|f(x) - f(y)| \leq c\|x - y\|$$

for all x and y, where $\|.\|$ is usually the l^2 norm in the $\mathbb{R}^n$ space. Suppose for each point x, the above inequality holds for all y in a neighbourhood of x, and then f is locally Lipschitz continuous. We refer the interested reader to Kloeden & Platen (1999) or Milstein (1995) for an overview on the classical numerical methods for SDEs. On the other hand, we shall notice that in general, to implement a Milstein scheme or a higher order Stochastic Taylor scheme for SDEs of multiple dimensions requires the simulation of some Itô integrals, called the Lévy area. The computation is generally not efficient, except in dimensional two, where we have, for instance, Gaines & Lyons (1994),Wiktorsson (2001) and Malham & Wiese (2014).

In the end, let us be more precise on the non-Lipschitz property of the Heston model. In the standard literature, such as Kloeden & Platen (1999) or Milstein

(1995), when we consider whether a multi-dimensional SDE has Lipschitz continuous coefficients, we take the SDE as a whole system. The property of Lipschitz continuity is independent of which numerical method to apply for simulation. Since the Heston model is a two-dimensional SDE, and the diffusion parts of both the variance process and the asset process involve the square root of the variance process, they are non-Lipschitz in terms of the variance process. Therefore, in this sense, it is fair to say that the Heston model has non-Lipschitz coefficients. However, in this thesis, we consider the numerical method that simulate the variance process exactly. We have the one-dimensional SDE

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2),$$

the coefficients of which are Lipschitz continuous in terms of the asset process, given the path of the variance process. In other words, the coefficients of the asset process are Lipschitz continuous only when the variance process is known. This does not mean that our analysis of the convergence rate falls into the scope of the classical analysis for SDEs with Lipschitz coefficients, because the convergence rate is not conditioned on the variance process.

2.3 Numerical Methods for the Heston Model

In this subsection, we shall provide a review of the literature on the numerical methods for the Heston model. Due to that the variance process can be simulated separately, we can first apply one time-discrete scheme for the variance process, then use another for the asset process based on the variance process approximated. The numerical performance, obviously, depends on the time discretization of the variance process, and that of the asset process.

We notice that although the variance process at a given time follows a non-central chi-square distribution, the exact simulation is generally quite computational inconvenient in practice. In the literature, these methods can be divided into the direct inversion methods and the acceptance-rejection methods. The general procedure of a direct inversion method is basically to first simulate a uniform random variable $U = u$ on $[0,1]$, and then to find a value x of X, which is the underlying random variable for simulation, such that the probability function satisfies $F(x) = u$. The calculation of x is generally very time-consuming, if there is no closed form of the inverse function $x = F^{-1}(u)$, particularly in pricing path-dependent options. An acceptance-rejection method, on the other hand, provides an alternative approach by first simulating a random variable Y that is easier to generate than the underlying random variable X we expect to simulate. A sample of Y can either be accepted as a sample of X or be rejected. The probability it is accepted is determined by the ratio of the densities of X and Y at the sample point. The methods include but are not limited to, Ahrens & Dieter (1974), Marsaglia & Tsang (2000) and Malham & Wiese (2013). However, in financial applications, the acceptance-rejection method is generally not favoured, due to that the number of samples required at each time step varies depending on the model inputs and state variables, which may introduce a large Monte Carlo bias into the sensitivity analysis, the analysis of the sensitivity of an option price to a certain parameter (Glasserman 2003, Chapter 7). For the same reason, the implementation of the quasi Monte Carlo simulation would become difficult, as discussed in a number of articles, for instance, Van Haastrecht & Pelsser (2010).

We start with a standard Euler scheme on the variance process. Let $\hat{V}_t$ be the time-discrete approximation of V_t at time $t \in [0,T]$, and we have

$$\hat{V}_{(i+1)h} = \hat{V}_{ih} + k(\theta - \hat{V}_{ih})h + \sqrt{\hat{V}_{ih}}\Delta W^1_{ih},$$

for $i = 0, 1, ..., T/h$, where $\Delta W^1_{ih} = W^1_{(i+1)h} - W^1_{ih}$, and $\hat{V}_0 = V_0$. As we can see from the standard Euler discretization above, given $\hat{V}_{ih}$, the distribution of $\hat{V}_{(i+1)h}$ is Normal, and thus there is a positive probability that the value of $\hat{V}_{(i+1)h}$ is negative, which is contrary to the property that the variance process is always non-negative. Further, the square root implies that such a Euler scheme would cause a serious computational failure when it is implemented in a computer. To avoid this problem, many authors have introduced the modified Euler schemes, which have a single general framework, given by Lord, Koekkoek & Van Dijk (2010),

$$\begin{aligned}\bar{V}_{(i+1)h} &= f_1(\bar{V}_{ih}) + k(\theta - f_2(\bar{V}_{ih}))h + \sqrt{f_3(\bar{V}_{ih})}\Delta W^1_{ih}\\ \hat{V}_{(i+1)h} &= f_3(\bar{V}_{(i+1)h}),\end{aligned}$$

where the functions $f_i(x) = x^+$, $|x|$, or x, $i = 1, 2, 3$, depending on the specific schemes, and $\bar{V}_0 = \hat{V}_0 = V_0$. The work includes Berkaoui et al. (2008), Bossy & Diop (2007), Deelstra & Delbaen (1998), Diop (2003), Higham & Mao (2005), and Lord et al. (2010), in all of which $\hat{V}_t$ at any time $t \in [0, T]$ is ensured to be non-negative. For instance, in Lord et al. (2010), they set

$$f_1(x) = x, \quad f_2(x) = x^+, \quad f_3(x) = x^+,$$

which they called the full truncation scheme. These methods mentioned, except for Higham & Mao (2005), simulate the asset process $(S_t)_{t\geq 0}$ in isolation. Lord et al. (2010) further proposed the logarithmic Euler discretization for the asset process of the form

$$\ln(\hat{S}_{(i+1)h}) = \ln(\hat{S}_{ih}) + (r - \frac{1}{2}\hat{V}_{ih})h + \sqrt{\hat{V}_{ih}}(\rho\Delta W^1_{ih} + \sqrt{1-\rho^2}\Delta W^2_{ih}),$$

where $\hat{S}_t$ is the time-discrete approximation of the asset process S_t at time $t \in [0, T]$, with $\hat{S}_0 = S_0$, and $\Delta W^2_{ih} = W^2_{(i+1)h} - W^2_{ih}$, independent of ΔW^1_{ih}. Here, ΔW^1_{ih} is the same as that in the Euler discretization of the variance process. After a numerical comparison of these methods for the Heston model of several sets of realistic parameters, Lord et al. (2010) claimed that their approach, the full truncation scheme, is the most efficient.

However, none of these methods provided a strong or a weak convergence rate for the Heston model, although the strong convergence was proved in Higham & Mao (2005) and Lord et al. (2010), and recently in Cozma & Reisinger (2015) for options of various payoff functions. The classical theory on the Euler scheme says when the drift and diffusion coefficients satisfy the global Lipschitz and linear growth conditions, then the standard strong order 1/2 and the weak order 1 are guaranteed. The strong order is retained by relaxing the global Lipschitz continuity to only local Lipschitz continuity, see Gyöngy (1998). Nevertheless, the variance process is not locally Lipschitz continuous in the neighbourhood of zero.

Kahl & Jäckel (2006) suggested an implicit Milstein scheme for the variance process, and a scheme for the logarithmic asset process, which they called the IKJ scheme. The discretization has the form

$$\begin{aligned}
\ln(\hat{S}_{(i+1)h}) &= \ln(\hat{S}_{ih}) + rh - \frac{1}{4}(\hat{V}_{(i+1)h}h + \hat{V}_{ih})h + \rho\sqrt{\hat{V}_{ih}}\Delta W^1_{ih} \\
&\quad + \frac{1}{2}\left(\sqrt{\hat{V}_{(i+1)h}} + \sqrt{\hat{V}_{ih}}\right)(\Delta W^2_{ih} - \rho\Delta W^1_{ih}) + \frac{1}{4}\sigma\rho((\Delta W^1_{ih})^2 - 1) \\
\hat{V}_{(i+1)h} &= \frac{\hat{V}_{ih} + k\theta h + \sigma\sqrt{\hat{V}_{ih}}\Delta W^1_{ih} + \frac{1}{4}\sigma^2((\Delta W^1_{ih})^2 - h)}{1 + kh},
\end{aligned}$$

where $\Delta W^1_{ih} = W^1_{(i+1)h} - W^1_{ih}$ and $\Delta W^2_{ih} = W^2_{(i+1)h} - W^2_{ih}$ as discussed, and they are mutually independent. Note that only when $4k\theta > \sigma^2$, the path of the variance process is ensured to be non-negative. Again, it was shown by Lord et al. (2010) that the full truncation scheme outperforms the IKJ scheme numerically. In our opinion, this result is not surprising. The standard Milstein scheme has the same weak order as the standard Euler scheme, and there seems to be no direct evidence that a numerical scheme with a higher strong order would reduce the mean square error if it does not improve the weak order, even for path-dependent option pricing. The mean square error, as discussed, is another important measure for the efficiency of a numerical scheme, particularly in the numerical test.

Alfonsi (2005) proposed an implicit Euler scheme, purely for the variance process, which is out of the Euler discretization form by Lord et al. (2010). This scheme has the standard strong and weak order, 1/2 and 1 respectively, but its application is limited to the parameter regime $4k\theta > \sigma^2$. To be more specific, let $Y_t = \sqrt{V_t}$, and we can write

$$\mathrm{d}Y_t = \left(\frac{\alpha}{Y_t} + \beta Y_t\right)\mathrm{d}t + \gamma \mathrm{d}W^1_t$$

with

$$\alpha = \frac{4k\sigma - \theta^2}{8}, \quad \beta = -\frac{k}{2}, \quad \gamma = \frac{\theta}{2}.$$

This transformation is known as the Lamperti transformation, which shifts the non-linearity from the diffusion coefficient into the drift coefficient. The drift coefficient $f(y) = \frac{\alpha}{y} + \beta y$ satisfies the one-sided Lipschitz condition, which is important to control the error of an implicit Euler scheme, see Higham et al. (2002). By implementing the Euler discretization on the SDE above, and then solving the equation, we obtain

$$\hat{Y}_{(i+1)h} = \frac{\hat{Y}_{(i+1)h} + \gamma \Delta W^1_{ih}}{2(1-\beta h)} + \sqrt{\left(\frac{\hat{Y}_{(i+1)h} + \gamma \Delta W^1_{ih}}{2(1-\beta h)}\right)^2 + \frac{\alpha h}{1-\beta h}},$$

with the initial $\hat{Y}_0 = \sqrt{V_0}$. We let $\hat{V}_{ih} = \hat{Y}^2_{ih}$, for all $i = 0, 1, ..., T/h$. Dereich et al. (2012) showed that with the piecewise linear interpolation, the strong order is 1/2 when $2k\theta > \sigma^2$, and Alfonsi (2013) demonstrated that without any interpolation, the strong order is 1 under a more restrictive condition $k\theta > \sigma^2$.

Recently, Alfonsi (2010) extended the method by Ninomiya & Victoir (2008) and presented two higher order schemes, one with the second weak order and the other with the third weak order, for the variance process. The author further extended the former for the Heston model, but there is no convergence rate obtained. These methods rely on the idea of scheme composition, the splitting of the differential operator. They are technically complicated, so we will skip the mathematical detail here. One advantage of the results is that they apply without any restriction on the model parameter and the theoretical weak order has been proved for the variance process with the smooth test functions whose derivatives are of polynomial growth. However, the author claimed there is no hope the analysis works for the Heston model, as it does not have the uniformly bounded moments.

At the same time, instead of time-discretizing the variance process, some authors considered either to approximate V_t at time t given $V_u, u < t$, with some random variables that can be easily simulated, or to simulate it almost exactly, as it has a non-central chi-square distribution. These methods include but are not limited to

Andersen (2008), Van Haastrecht & Pelsser (2010) and Malham & Wiese (2013).

Andersen (2008) made a breakthrough by approximating the variance process with the quadratic-exponential (QE) scheme, where the coefficients are determined by matching the first and second moments of the underlying random variables with those of the variance process. Specifically, let

$$m = \mathbb{E}(\hat{V}_{t+h}|\hat{V}_t) = \theta + (\hat{V}_t - \theta)e^{-kh}$$

$$s^2 = \mathrm{Var}(\hat{V}_{t+h}|\hat{V}_t) = \frac{\hat{V}_t\sigma^2 e^{-kh}}{k}(1-e^{-kh}) + \frac{\theta\sigma^2}{2k}(1-e^{-kh})^2,$$

and $\psi = \frac{s^2}{m^2}$. If $\psi \leq 1.5$, the author wrote $\hat{V}_{t+h}$ as

$$\hat{V}_{t+h} = a(b + Z_V)^2,$$

where Z_V is a standard Normal random variable, and

$$b^2 = 2\psi^{-1} - 1 + \sqrt{2\psi^{-1}}\sqrt{2\psi^{-1} - 1}$$

$$a = \frac{m}{1+b^2}.$$

On the other hand, if $\psi > 1.5$, then the author approximated the probability by

$$\mathbb{P}(\hat{V}_{t+h} \in [x, x+h]) \approx (p\delta(0) + \beta(1-p)e^{-\beta x})\mathrm{d}x,$$

where $\delta(.)$ is a Dirac delta-function, and p, β are constant to be determined. We can get the approximated cumulative distribution function $\Psi(.)$ of $\hat{V}_{t+h}$, after the integration of the formula above, and it follows that

$$\hat{V}_{t+h} = \Psi^{-1}(U_V; \beta, p),$$

with U_V an uniform random variable on $[0,1]$, and $\Psi^{-1}(U_V = u) = 0$, if $0 \leq u \leq p$, and otherwise, $\Psi^{-1}(U_V = u) = \beta^{-1}\ln\left(\frac{1-p}{1-u}\right)$. The values of the coefficients p and β

are

$$p = \frac{\psi - 1}{\psi + 1} \quad and \quad \beta = \frac{2}{m(\psi + 1)}.$$

We comment that although Andersen's QE scheme contains a small bias due to that it matches only the first two moments of the underlying variance process, the method applies in all parameter regimes, and this, together with the stochastic trapezoidal rule to approximate the time integral of the variance process in the SDE of the logarithmic asset process, is highly efficient for the Heston model. The numerical tests in Andersen (2008) showed that it is significantly better than Kahl & Jäckel (2006) and Lord et al. (2010).

Van Haastrecht & Pelsser (2010) implemented an efficient caching technique for the variance process. To be more precise, they created a cache of the values of the inverse of the non-central chi-squared distribution function by means of conditioning on a truncated range of Poisson-values and pre-computing the corresponding chi-squared distribution functions, which requires some interpolation technique, such as the monotone cubic Hermite spline interpolation. Malham & Wiese (2013) proved the variance process can be represented by the sum of some generalized Gaussian random variables, for which they provided a direct inversion scheme. Specifically, let $\chi_d^2(\lambda)$ be a non-central chi-square random variable with degree of freedom d and non-centrality λ, and we can write $\chi_d^2(\lambda) = \chi_{d+2N}^2$, where N here is a Poisson random variable with mean $\lambda/2$, and χ_{d+2N}^2 is a chi-square random variable with degree of freedom $d + 2N$. They showed that for any positive integers p and q, there is a representation

$$\chi_{p/q}^2 = \sum_{i=1}^{p} |X_i|^{2q},$$

where X are independent generalized Gaussian random variables $N(0, 1, 2q)$. The almost exact simulation of X by direct inversion requires the Padé and Chebychev approximation of the cumulative distribution function of X. Both methods are almost exact, and apply without any restriction on the parameters. Their numerical results for the Heston model were demonstrated to be comparable to Andersen (2008), together with the same stochastic trapezoidal rule to discretize the SDE of the asset process. However, there is little research to explore the mathematical reason behind.

For the stochastic trapezoidal rule on the asset price, recall that from the Heston model, let $X_t = \ln(e^{-rt}S_t)$, and we can write

$$X_{t+h} = X_t + \frac{\rho}{\sigma}(V_{t+h} - V_t - k\theta h) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_t^{t+h} V_s \mathrm{d}s + \sqrt{1-\rho^2}\sqrt{\int_t^{t+h} V_s \mathrm{d}s}N \tag{2.3}$$

for any positive h, where N is a standard Normal random variable, independent of the variance process. Andersen (2008), Van Haastrecht & Pelsser (2010) and Malham & Wiese (2013) approximate the integral

$$\int_t^{t+h} V_s \mathrm{d}s \approx \frac{V_t + V_{t+h}}{2}h,$$

which is the stochastic trapezoidal rule. For all the methods we have discussed above, none of them provides a theoretical strong or weak convergence rate with respect to the Heston model. We shall show in this thesis that, by assuming the variance process is simulated exactly, it is possible to get an analytical weak order two, that is consistent with the standard rate of the stochastic trapezoidal rule, although the coefficient of the Heston model does not satisfy the global Lipschitz condition. Further, under the same assumption, we implement the Multi-level Monte Carlo on the Heston model, with the analytical strong convergence rate derived for the Multi-level Monte Carlo estimator we define.

To complete the simulation review, we shall notice that for a path-independent option, where the value of an option is only determined by the asset price at the maturity, it is not necessary to simulate the asset process path-wisely. Indeed, there are several methods in the literature that simulate directly the asset process at the final time. Broadie & Kaya (2006) wrote

$$X_T = X_0 + \frac{\rho}{\sigma}(V_T - V_0 - k\theta T) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_0^{T} V_s \mathrm{d}s + \sqrt{1-\rho^2}\sqrt{\int_0^{T} V_s \mathrm{d}s}N,$$

where N is a standard Normal random variable, independent of the variance process. The formula is the same as equation (2.3) for $t = 0$ and $h = T$. We notice that to simulate X_T, we have to simulate the pair

$$\left(V_T, \int_0^T V_s \mathrm{d}s \right).$$

Broadie & Kaya (2006) considered first to simulate V_T exactly, and then to simulate the integral $\int_0^T V_s \mathrm{d}s$ conditioned on V_T. The simulation is by the direct inversion of the probability distribution function obtained by the Fourier transform, as the characteristic function of $\int_0^T V_s \mathrm{d}s$ given V_T has been derived explicitly. However, their method has been criticised to be computational slow, and is thus not generally appealing in practice. There are two reasons for the computational inconvenience. First, the characteristic function changes as the value of V_T changes, which makes it impossible to pre-cache the distribution function. Second, the characteristic function contains a modified Bessel function of the first kind, the calculation of which is time-consuming. In the spirit of Broadie & Kaya (2006), Glasserman & Kim (2011) derived a representation

$$\left(\int_0^T V_s \mathrm{d}s \,\middle|\, V_0, V_T \right) = Y_1 + Y_2 + \sum_{j=1}^{\eta} Z_j,$$

where Y_1, Y_2, Z and η are mutually independent. The derivation was based on the decomposition of Bessel Bridges, as discussed in Pitman & Yor (1982). The specific representation are not shown here, but Y_1, Y_2 and Z can be written as the summation of some Gamma random variables, and η is a Bessel random variable. Only Y_1 and η rely on V_T while the other random variables Y_2 and Z are independent of V_T. This means we can tabulate the distributions of Y_2 and Z at the start of the simulation and then draw samples as needed by sampling from the table. Since the characteristic function of Y_1 is only composed of some exponential and hyperbolic functions, the computation of which is very easy when applying a direct inversion technique. Chan & Joshi (2013) developed this approach by providing two new methods to simulate V_T at the maturity T. For more discussions on Monte Carlo methods for the Heston model and for more general affine processes, we recommend the reader to Alfonsi (2015) for a review.

Finally, the density function of the logarithmic asset process has recently been formulated in del Baño Rollin et al. (2010), which is an infinite convolution of Bessel type densities, and is of $\mathcal{C}^{\infty}$. This finding is interesting as one may sample the logarithmic asset process directly from the inverse of its distribution function, which might be approximated, in a similar approach as in Malham & Wiese (2013) for the variance process, in order to avoid the time-consuming root-finding as in the usual approach for the inverse problem. Further, the scheme is unbiased without any discretization on the SDEs. To achieve it, we may have to establish a more specific link between the density of the logarithmic asset process and that of the variance process.

Chapter 3

Weak Convergence Rate for the Heston Model

3.1 Introduction

In this chapter, we derive the weak convergence rate of a time-discrete scheme for the Heston stochastic volatility model, that employs the stochastic trapezoidal rule to approximate the time integral of the variance process in the SDE of the logarithmic asset process, and simulates the variance process exactly. The test function we consider for the error criterion can be any polynomials of the logarithmic asset process. We show that the analytical weak convergence rate is two in all parameter regimes. The result is consistent with the standard rate of the stochastic trapezoidal rule, although the coefficient of the model does not satisfy the global Lipschitz condition. Finally, we extend the result for the SVJ Model, also with the second order weak convergence under the full parameter regime.

The stochastic trapezoidal discretization is widely used in the literature for numerical solutions of SDEs, although the theory seems not well-established, at least compared with the Euler discretization. Generally, the stochastic trapezoidal discretization has a higher weak convergence rate than the Euler discretization, and there is more flexibility to modify a trapezoidal discretization for a typical SDE. The traditional trapezoidal methods can be found in Kloeden & Platen (1999), Chapter 15, as weak-order-two explicit methods or predictor-corrector methods. The weak order two can be observed when these schemes are applied for SDEs with coefficients satisfying the usual conditions. However, these methods generally fail when applied to the Heston model. In terms of the second order numerical methods for the Heston model, recently, Ninomiya & Victoir (2008) presented a novel weak approximation with order two for non-Lipschitz SDEs, and they applied it to price Asian options under the Heston model. Alfonsi (2010) developed this approach by extending the method to the full parameter regime of the Heston model, and the author showed numerically that it converges weakly with order two. To the best of our knowledge, there seems to be no higher order algorithm in the literature for the Heston model.

3.2 Weak Convergence Rate of Stochastic Trapezoidal Rule

In this section, we shall develop the theory of stochastic trapezoidal rule, that are useful to the convergence analysis for the Heston model.

3.2.1 Problem Formulation

Consider the probability space $(\Omega, \mathcal{F}, P)$ with the filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ satisfying the usual conditions, where T is a constant. Let $(\phi_t, \varphi_t)_{0\leq t\leq T}$ be a pair of square-integrable stochastic processes on this probability space, such that one can sample exactly from their distributions. For any positive integer m, suppose the expectation below exists and is finite, we let the function

$$f_m(t_1, ..., t_m) = \mathbb{E}(\phi_{t_1}\phi_{t_2}...\phi_{t_m}\varphi_T)$$

on the simplex domain

$$\Omega_m = \{(t_1, t_2, ..., t_m)|0 \leq t_1 \leq t_2 \leq ... \leq t_m \leq T\}.$$

For notational convenience, we denote the stochastic trapezoidal approximation

$$\fint_0^T \phi_t^h \mathrm{d}t \triangleq \sum_{i=1}^{T/h} \frac{\phi_{h(i-1)} + \phi_{hi}}{2} h$$

with the step size h. Our goal in this section is to prove

$$\mathbb{E}\left[\varphi_T \left(\fint_0^T \phi_t^h \mathrm{d}t\right)^m\right] = \mathbb{E}\left[\varphi_T \left(\int_0^T \phi_t \mathrm{d}t\right)^m\right] + O(h^2) \tag{3.1}$$

for any m, under the following two assumptions.

Assumption 3.2.1. *The second order derivative*

$$\frac{\partial^2 f_m}{\partial t_j^2} \tag{3.2}$$

exists and is continuous on Ω_m for all $j = 1, 2, ..., m$.

Assumption 3.2.2. *The integral over the domain Ω_m is finite, i.e*

$$\int_{\Omega_m} \mathbb{E}|\phi_{t_1}\phi_{t_2}, ..., \phi_{t_m}\varphi_T| < \infty. \tag{3.3}$$

For convenience, we shall state the first the twice continuous differentiability assumption, and the second the integrability assumption. We emphasize that the function f_m is defined on the simplex domain Ω_m rather than on the cube $[0, T]^m$. Note that if the twice continuous differentiability assumption is satisfied for f_m on $[0, T]^m$, then it is in particular satisfied for f_m on Ω_m. However, the inverse is not true for general stochastic processes $(\phi_t)_{t \geq 0}$ and $(\varphi_t)_{t \geq 0}$, as there may be singularities on $[0, T]^m$. There is an example in Remark 3.2.1, where f_m is only twice continuously differentiable on Ω_m, instead of on $[0, T]^m$.

Remark 3.2.1. *In assumption 3.2, we impose the twice continuous differentiability on the simplex domain rather than on the cube $[0, T]^m$, because the latter is restrictive. For example, we consider $\phi_t = W_t$ and $\varphi_T = 1$. It follows that $\mathbb{E}(W_t W_s) = t$, if $t \leqslant s$, and $\mathbb{E}(W_t W_s) = s$, if $t \geqslant s$. Then, the first order partial derivative with respect to t is 1, if $t \leqslant s$, and it is 0, if $t \geqslant s$. This indicates that the partial* *derivative is not well-defined on the boundary $t = s$, due to the left derivative is not the same as the right one.*

Remark 3.2.2. *If the stochastic processes can be written as $\phi_t = a(t, W_t)$, and $\varphi_T = b(T, W_T)$, where $(W_t)_{0\leq t\leq T}$ is a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$, and $a(t, x)$, $b(t, x) : \mathbb{R}^+ \times \mathbb{R} \to \mathbb{R}$ are certain functions, then our problem is reduced to the evaluation of Wiener space integral (Kloeden & Platen 1999, Section 17.1).*

3.2.2 Problem Transferred to a Deterministic Analysis

To establish the theorem, we shall transfer the stochastic problem to a deterministic problem. We shall begin with the convergence analysis of (3.1) for the first moment with $m = 1$, and then extend the result for higher moments with $m \geq 2$.

Lemma 3.2.1. *Let $f(t) = \mathbb{E}(\phi_t \varphi_T)$. If $f(t) \in C^2[0, T]$, and $\int_0^T \mathbb{E}|\phi_t \varphi_T| dt < \infty$, then we have*

$$\mathbb{E}\left(\varphi_T \fint_0^T \phi_t^h \mathrm{d}t\right) = \mathbb{E}\left(\varphi_T \int_0^T \phi_t \mathrm{d}t\right) + O(h^2).$$

Proof. As $f(t) \in C^2[0, T]$, from the result of deterministic trapezoidal rule on $\int_0^T f(t)\mathrm{d}t$, we obtain

$$\begin{aligned}
\mathbb{E}\left(\varphi_T \fint_0^T \phi_t^h \mathrm{d}t\right) &= \sum_{i=1}^{T/h} \frac{\mathbb{E}(\phi_{h(i-1)}\varphi_T) + \mathbb{E}(\phi_{hi}\varphi_T)}{2} h \\
&= \int_0^T \mathbb{E}(\phi_t \varphi_T) dt + O(h^2) \\
&= \mathbb{E}\left(\varphi_T \int_0^T \phi_t \mathrm{d}t\right) + O(h^2),
\end{aligned}$$

where the exchange of the order of the expectation and the integral is justified by the Fubini theorem, under the condition $\int_0^T \mathbb{E}|\phi_t \varphi_T| \mathrm{d}t < \infty$. □

Remark 3.2.3. *Lemma 3.2.1 can be generalized to the case where $\int_0^T \phi_t \mathrm{d}t$ is approximated by any quadrature rule, and then the convergence rate would be the same as that of the corresponding deterministic rule under the similar conditions. For example, if we apply the composite Simpson's rule, then we have $O(h^4)$. However,* *for the generalized stochastic quadrature rule, the extension of analysis to higher moments is difficult. We shall focus on the stochastic trapezoidal rule in this thesis.*

We notice that the idea implicated in Lemma 3.2.1 is to transfer the problem to the trapezoidal rule on the integral of the deterministic function $f(t) = \mathbb{E}(\phi_t \varphi_T)$. This idea can be extended for higher moments with $m \geq 2$, which corresponds to a multi-variable deterministic function $\mathbb{E}(\phi_{t_1}\phi_{t_2}...\phi_{t_m}\varphi_T)$. Specifically, under the integrability assumption (3.3), we have

$$\mathbb{E}\left[\varphi_T \left(\int_0^T \phi_t \mathrm{d}t\right)^m\right] = \int_0^T \int_0^T ... \int_0^T \mathbb{E}(\phi_{t_1}\phi_{t_2}...\phi_{t_m}\varphi_T)\mathrm{d}t_1\mathrm{d}t_2...\mathrm{d}t_m, \tag{3.4}$$

and we consider to apply the trapezoidal rule on the right side of (3.4). For notational simplicity, we let $\bar{f}_m$ to be the extension of the function f_m with the same formula

$$\bar{f}_m(t_1, ..., t_m) = \mathbb{E}(\phi_{t_1}\phi_{t_2}...\phi_{t_m}\varphi_T),$$

but on the cube

$$\bar{\Omega}_m = \{(t_1, t_2, ..., t_m)|0 \leq t_1, t_2, ..., t_m \leq T\} \triangleq [0, T]^m.$$

Then, the right side of (3.4) can be written as $\int_{\bar{\Omega}_m} \bar{f}_m$.

If we are capable of proving the trapezoidal rule on $\int_{\bar{\Omega}_m} \bar{f}_m$ is second order accurate, under the condition that f_m is twice continuously differentiable on Ω_m, then it is not difficult to verify (3.1). To be more specific, by the trapezoidal rule, we can write

$$\begin{aligned}
\int_{\bar{\Omega}_m} \bar{f}_m &= \int_0^T \int_0^T ... \int_0^T \bar{f}_m(t_1, t_2, ..., t_m)\mathrm{d}t_1\mathrm{d}t_2...\mathrm{d}t_m \\
&\approx \sum_{j_1=0}^{T/h} \sum_{j_2=0}^{T/h} ... \sum_{j_m=0}^{T/h} a_{j_1,h}a_{j_2,h}...a_{j_m,h}\bar{f}_m(j_1 h, j_2 h, ..., j_m h) \\
&\triangleq \mathcal{T}\bar{f}_m,
\end{aligned}$$

where

$$a_{j,h} = \begin{cases} \frac{1}{2}h, & j = 0, \frac{T}{h}, \\ h, & else. \end{cases}$$

and $\mathcal{T}$ is the trapezoidal rule. We expect to prove that

$$\left| \int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m \right| = O(h^2),$$

under the condition that f_m is twice continuously differentiable on Ω_m. However, the standard theory of trapezoidal rule to guarantee the second order convergence requires a twice continuously differentiable integrand on a cube, and we see that $\bar{f}_m$ is only twice continuously differentiable on each simplex component of $\bar{\Omega}_m$. The extension of analysis is not trivial, although there is a strong link in between.

3.2.3 Preliminary Result

For the standard analysis, there is an error bounded due to Haber (1970). The majority of notations here are inherited from the original paper. Consider an integral on the cube $[0,1]^m$ with the integrand $g : [0,1]^m \to \mathbb{R}$. Suppose $Q_1, Q_2, ..., Q_m$ are quadrature formulas for the interval $[0,1]$, we can write

$$\begin{aligned} \int_{[0,1]^m} g &= \int_0^1 ... \int_0^1 g(x^1, x^2, ..., x^m) \mathrm{d}x^1 \mathrm{d}x^2 ... \mathrm{d}x^m \\ &\approx \int_0^1 ... \int_0^1 Q_1(g; x^1) \mathrm{d}x^2 \mathrm{d}x^3 ... \mathrm{d}x^m \\ &\approx \int_0^1 ... \int_0^1 Q_2(Q_1(g; x^1); x^2) \mathrm{d}x^3 ... \mathrm{d}x^m \\ &\approx ... \\ &\approx Q_m(Q_{m-1}(...Q_1(g; x^1); x^2)...; x^m), \end{aligned} \tag{3.5}$$

where

$$Q_i(g) = \sum_{j=1}^{n_i} a_{j,i} g(x_{j,i}).$$

for $i = 1, 2, ..., m$. Then, the expression (3.5) is

$$\sum_{j_1=1}^{n_1} \sum_{j_2=1}^{n_2} ... \sum_{j_m=1}^{n_m} a_{j_1,1} a_{j_2,2} ... a_{j_m,m} g(x_{j_1,1}, x_{j_2,2}, ..., x_{j_m,m}),$$

and we denote it by

$$\left(\prod_{i=1}^{m} Q_i \right) g.$$

There is an error estimate for such formulas. Let us say we know that

$$\left| \int_0^1 g(x^1, ..., x^m) \mathrm{d}x^i - Q_i(g; x^i) \right| \leq \mathrm{E}_i, \tag{3.6}$$

for all values of $x^1, ..., x^{i-1}, x^{i+1}, ..., x^m$ lying between 0 and 1, and we have such an error estimate for each $i = 1, 2, ..., m$. Then, we have

$$\left| \int_{[0,1]^m} g - \left(\prod_{i=1}^{m} Q_i \right) g \right| \leq \mathrm{E}_1 + \mathrm{A}_1 \mathrm{E}_2 + \mathrm{A}_1 \mathrm{A}_2 \mathrm{E}_3 + ... + \mathrm{A}_1 \mathrm{A}_2 ... \mathrm{A}_{m-1} \mathrm{E}_m, \tag{3.7}$$

where A_i denotes the sum of the absolute values of the coefficients in the formula Q_i.

One sufficient condition for the inequality (3.6) to be satisfied is that the integrand g is smooth on the cube $[0, 1]^m$, and all the partial derivatives are bounded. In the case of trapezoidal rule with $Q_1 = Q_2 = ... = Q_m$ of step size h, suppose

$$\left| \frac{\partial^2 g}{(\partial x^i)^2} \right| \leq M$$

for all $i = 1, 2, ..., m$ throughout $[0, 1]^m$, where M is a constant, it is not difficult to see that there exists a constant K, such that

$$\left| \int_0^1 g(x^1, ..., x^m) \mathrm{d}x^i - Q_i(g; x^i) \right| \leq K h^2,$$

for all i. Since $\mathrm{A}_i = 1$, it follows by (3.6) and (3.7) that

$$\left| \int_{[0,1]^m} g - \left(\prod_{i=1}^{m} Q_i \right) g \right| \leq K h^2,$$

and that is how the second order convergence is derived.

However, in our problem, we assume the integrand g is twice continuously differentiable on a simplex domain instead of a cube, and then (3.6) is only satisfied for all $x^1, ..., x^{i-1}, x^{i+1}, ..., x^m$ taking values on the grids $\{0, h, 2h, ..., 1\}$, rather than all these taking values in $[0, 1]$.

On the other hand, for smooth integral over a simplex domain, the simplest approach may be to first transfer the simplex domain to the cube by changing the variables of integration with the aid of the Jacobian matrix, and then apply a quadrature formula to approximate the integral through the Cartesian product. If the integrand is smooth on the simplex domain, its transformation is also smooth on the cube. Then, there is a classical theorem by Haber (1970) to guarantee the desired convergence order of the quadrature rule applied, as discussed. It is also notable that this approach can be applicable even though there is a singularity on the integrand (Duffy 1982). In our problem, the trapezoidal rule is defined on the simplex domain with the equidistant step size. If we insist to transfer the domain to a cube, we will find that the step size on the new domain is irregular, and it is impossible to apply the Cartesian product. This makes the analysis rather difficult. In terms of those methods for integration over the simplex without changing the domain, there are a class of standard cubature rules, quadrature in dimensions three and higher, proposed by Hammer et al. (1956). They are specially tailored for the simplex domain, and are constructed in such a way that they produce no-error approximations when the integrand is a polynomial of a given degree. However, the approximations are based on a few fixed points on the domain, and thus have little to do with the trapezoidal discretization we consider with fixed step size, but with non-fixed points. The reader can visit Davis & Rabinowitz (1984), Chapter 5 for a review of quadrature rules on multiple integrals, or visit Pond (2010) for a more detailed discussion on theory of integration over the simplex. Generally, these standard techniques mentioned above are powerful to approximate the value of the integral, if the integrand is smooth. However, the trapezoidal rule we consider is not designed for approximating a numerical integral, but rather for analysing a numerical problem arising in SDEs. It seems difficult to find a directly relevant theory in the literature. One possible reason is that for the smooth integrand, it may suffice to use the standard techniques, if we are only interested in how to approximate the integral efficiently.

3.2.4 Proof for Theorem

In the following part, we attempt to provide a proof that

$$\left| \int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m \right| = O(h^2),$$

under the assumption that f_m is twice continuously differentiable on Ω_m. The central step of the proof is to understand the structure of a representation, which is an expansion of a multiple integral. In our analysis, we decompose

$$\begin{aligned}
&\left| \int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m \right| \\
\leqslant &\left| \int_{\bar{\Omega}_m} \bar{f}_m - \sum_{j_1=0}^{T/h} a_{j_1,h} \int_0^T ... \int_0^T \bar{f}_m(j_1 h, t_2, ..., t_m) \mathrm{d}t_2 ... \mathrm{d}t_m \right| \\
&+ \sum_{j_1=0}^{T/h} a_{j_1,h} \left| \int_0^T ... \int_0^T \bar{f}_m(j_1 h, t_2, ..., t_m) \mathrm{d}t_2 ... \mathrm{d}t_m \right. \\
&\quad \left. - \sum_{j_2=0}^{T/h} a_{j_2,h} \int_0^T ... \int_0^T \bar{f}_m(j_1 h, j_2 h, t_3, ..., t_m) \mathrm{d}t_3 ... \mathrm{d}t_m \right| \\
&+ \sum_{j_1=0}^{T/h} \sum_{j_2=0}^{T/h} a_{j_1,h} a_{j_2,h} \left| \int_0^T ... \int_0^T \bar{f}_m(j_1 h, j_2 h, t_3, ..., t_m) \mathrm{d}t_3 ... \mathrm{d}t_m \right. \\
&\quad \left. - \sum_{j_3=0}^{T/h} a_{j_3,h} \int_0^T ... \int_0^T \bar{f}_m(j_1 h, ..., j_3 h, t_4, ..., t_m) \mathrm{d}t_4 ... \mathrm{d}t_m \right| \\
&+ ... \\
&+ \sum_{j_1=0}^{T/h} ... \sum_{j_{m-1}=0}^{T/h} a_{j_1,h} ... a_{j_{m-1},h} \left| \int_0^T \bar{f}_m(j_1 h, ..., j_{m-1} h, t_m) \mathrm{d}t_m \right. \\
&\quad \left. - \sum_{j_m=0}^{T/h} a_{j_m,h} \bar{f}_m(j_1 h, ..., j_m h) \right|,
\end{aligned} \tag{3.8}$$

Note that in (3.8), we have by definition that

$$\sum_{j_1=0}^{T/h} \dots \sum_{j_i=0}^{T/h} a_{j_1,h} \dots a_{j_i,h} = T^i, \tag{3.9}$$

for all $i = 1, 2, .., m-1$, where T^i is a constant, with fixed i. In addition, the expression in each $|.|$ of (3.8) can be regarded as the global error of the trapezoidal rule, and we intend to prove that it is of order two, regardless of the values of j_i, $i = 1, ..., m$. Before we derive the global error, we shall look at the local error, the analysis of which relies on the structure of a representation, which is the summation of some integrals.

Let us set up the notations first. Let the vector

$$\tau_i = (j_1 h, j_2 h, ..., j_i h)$$

for all $i = 1, 2, ..., m-1$, where $j_1, j_2, ..., j_i \in \{0, 1, 2, ..., T/h\}$, and let τ_0 be a null vector. In our analysis, we assume τ_i is given and we denote the function

$$I_{\tau_{m-1}}(t_m) = \bar{f}_m(\tau_{m-1}, t_m)$$

and

$$I_{\tau_i}(t_{i+1}) = \int_0^T \dots \int_0^T \bar{f}_m(\tau_i, t_{i+1}, ..., t_m) \mathrm{d}t_{i+2} \dots \mathrm{d}t_m \triangleq \int_{\bar{\Omega}_{\tau_i^*}} \bar{f}_m$$

for $i = 0, 1, ..., m-2$, where $I_{\tau_{m-1}}(t_m)$ and $I_{\tau_i}(t_{i+1})$ are similar to those integrals in $|.|$ of (3.8). Here,

$$\tau_i^* = (\tau_i, t_{i+1}),$$

and integration domain

$$\bar{\Omega}_{\tau_i^*} = \{(\tau_i^*, t_{i+2}, ..., t_m) | t_{i+2}, ..., t_m \in [0, T]\}$$

is a subset of $\bar{\Omega}_m$. Since $\bar{f}_m$ is not twice continuously differentiable everywhere on $\bar{\Omega}_m$, the analysis requires to divide $\bar{\Omega}_m$ into several pieces, on each of which $\bar{f}_m$ is twice continuously differentiable. This decomposition depends on τ_i^*. In addition, for any $l = 1, 2, ..., T/h$, we define G_{l,τ_i} as the integral of $I_{\tau_i}(t_{i+1})$ over $[(l-1)h, lh]$, and we have

$$G_{l,\tau_i} \triangleq \int_{(l-1)h}^{lh} I_{\tau_i}(t_{i+1})\mathrm{d}t_{i+1},$$

where $i = 0, 1, ..., m-1$. Furthermore, we let Q_{l,τ_i} be the trapezoidal approximation on G_{l,τ_i} with the step size h, and we have

$$Q_{l,\tau_i} \triangleq \frac{h}{2}\left[I_{\tau_i}\left((l-1)h\right) + I_{\tau_i}\left(lh\right)\right].$$

In order to demonstrate that each expression in $|.|$ of (3.8) is of order two, which is the global error, the key step is to prove the local error

$$|G_{l,\tau_i} - Q_{l,\tau_i}| \leq L_i h^3$$

for any l and values of τ_i given, where L_i is a constant with i fixed. By the well-known theorem of trapezoidal rule and the definitions of $I_{\tau_i}(t_{i+1})$, G_{l,τ_i} and Q_{l,τ_i}, we shall firstly show that

$$I_{\tau_i}(t_{i+1}) \in C^2[(l-1)h, lh] \tag{3.10}$$

for any $i = 0, 1, ..., m-1$. The proof of (3.10) is trivial when $i = m-1$. For $i = 0, ..., m-2$, the integrand $\bar{f}_m$ of $I_{\tau_i}(t_{i+1})$ is not twice continuously differentiable everywhere on $\bar{\Omega}_{\tau_i^*}$, and thus, it is more convenient for us to split $\bar{\Omega}_{\tau_i^*}$ into several sub-domains, denoted as $D_{\tau_i^*}$, such that for each of these sub-domains, $\bar{f}_m$ is not twice continuously differentiable only on its boundary. In other words, we separate $\bar{\Omega}_{\tau_i^*}$ from the place where $\bar{f}_m$ is not twice continuously differentiable. Therefore, we have the following representation

$$I_{\tau_i}(t_{i+1}) = \int_{\Omega_{\tau_i^*}} \bar{f}_m = \sum_{\{D_{\tau_i^*}\}} \int_{D_{\tau_i^*}} f_m \tag{3.11}$$

for $i = 0, 1, ..., m-2$, where $\{D_{\tau_i^*}\}$ denotes the set containing all $D_{\tau_i^*}$. Here, due to the twice continuously differentiability of the integrand $\bar{f}_m$ on $D_{\tau_i^*}$, we can write $\bar{f}_m$ as f_m by simply switching the order of their variables. Although the representation above can be formulated analytically, the exact form is not essential for our proof. However, we shall give some examples later so the reader can have a more clear picture on how we split $D_{\tau_i^*}$ and which mathematical structure $\int_{D_{\tau_i^*}} f_m$ has. Recall that $\tau_i^* = (\tau_i, t_{i+1})$, in which τ_i is fixed, and t_{i+1} is a variable. This means the sub-domain $D_{\tau_i^*}$ changes according to the variable $t_{i+1} \in [(l-1)h, lh]$. Nevertheless, the analytical formula of the representation is always the same when t_{i+1} varies in $[(l-1)h, lh]$. This property is of vital importance to ensure $I_{\tau_i}(t_{i+1})$ is twice continuously differentiable for all $t_{i+1} \in [(l-1)h, lh]$.

Example 3.2.1. *We set $m = 3$, where m is the number of variables in $\bar{f}_m$ or f_m. Suppose $\tau_1 = (j_1 h)$ and $\tau_1^* = (\tau_1, t_2)$, where $j_1 h \leqslant t_2$. We have*

$$
\begin{aligned}
I_{\tau_1}(t_2) &= \int_0^T \bar{f}_3(j_1 h, t_2, t_3) \mathrm{d}t_3 \\
&= \int_0^{j_1 h} f_3(t_3, j_1 h, t_2) \mathrm{d}t_3 + \int_{j_1 h}^{t_2} f_3(j_1 h, t_3, t_2) \mathrm{d}t_3 + \int_{t_2}^{T} f_3(j_1 h, t_2, t_3) \mathrm{d}t_3,
\end{aligned}
$$

where $\bar{f}_3$ is replaced by f_3 in the second equation by switching the order of variables $j_1 h$, t_2 and t_3, according to their magnitude order.

Example 3.2.2. *Suppose $m = 4$, and $\tau_2 = (j_1 h, j_2 h)$ and $\tau_2^* = (\tau_2, t_3)$, where $j_1 h \leqslant j_2 h \leqslant t_3$, then we have*

$$
\begin{aligned}
I_{\tau_2}(t_3) &= \int_0^T \bar{f}_4(j_1 h, j_2 h, t_3, t_4) \mathrm{d}t_4 \\
&= \int_0^{j_1 h} f_4(t_4, j_1 h, j_2 h, t_3) \mathrm{d}t_4 + \int_{j_1 h}^{j_2 h} f_4(j_1 h, t_4, j_2 h, t_3) \mathrm{d}t_4 \\
&\quad + \int_{j_2 h}^{t_3} f_4(j_1 h, j_2 h, t_4, t_3) \mathrm{d}t_4 + \int_{t_3}^{T} f_4(j_1 h, j_2 h, t_3, t_4) \mathrm{d}t_4,
\end{aligned}
$$

where again $\bar{f}_4$ is replaced by f_4 in the second equation by switching the order of variables $j_1 h$, $j_2 h$, t_3 and t_4, in such way that the smallest comes first, the second smallest comes the second, and so on.

Example 3.2.3. *Let $m = 4$, $\tau_1 = (j_1 h)$ and $\tau_1^* = (\tau_1, t_2)$, where we also assume*

$j_1h \leqslant t_2$. Then, we obtain

$$\begin{aligned}
I_{\tau_1}(t_2) &= \int_0^T \int_0^T \bar{f}_4(j_1h, t_2, t_3, t_4)\mathrm{d}t_3\mathrm{d}t_4 \\
&= 2\Bigg(\int_0^{j_1h}\int_0^{t_4} f_4(t_3, t_4, j_1h, t_2)\mathrm{d}t_3\mathrm{d}t_4 + \int_{j_1h}^{t_2}\int_0^{j_1h} f_4(t_3, j_1h, t_4, t_2)\mathrm{d}t_3\mathrm{d}t_4 \\
&\quad + \int_{t_2}^{T}\int_0^{j_1h} f_4(t_3, j_1h, t_2, t_4)\mathrm{d}t_3\mathrm{d}t_4 + \int_{j_1h}^{t_2}\int_{j_1h}^{t_4} f_4(j_1h, t_3, t_4, t_2)\mathrm{d}t_3\mathrm{d}t_4 \\
&\quad + \int_{t_2}^{T}\int_{j_1h}^{t_2} f_4(j_1h, t_3, t_2, t_4)\mathrm{d}t_3\mathrm{d}t_4 + \int_{t_2}^{T}\int_{t_2}^{t_4} f_4(j_1h, t_2, t_3, t_4)\mathrm{d}t_3\mathrm{d}t_4\Bigg),
\end{aligned}$$

in which the integrands f_4 of the integrals on the right side have different combinations of the variables j_1h, t_2, t_3 and t_4. Since $j_1h \leqslant t_2$, we have $\frac{4!}{2!} = 12$ such integrals. However, an integral on the domain with $t_3 \leq t_4$ takes the same value as that on the corresponding domain with $t_4 \leq t_3$, due to symmetry. This is why we put coefficient 2 here, with only $\frac{4!}{2!2!} = 6$ integrals.

We notice that in the examples above, if we instead let $t_2 \leqslant j_1h$ in the first and the third examples, then their analytical formulas would change accordingly. Generally, if the magnitude order of the elements in τ_i^* keeps the same, then the analytical formula of the representation would not change. To be more specific, recall that $\tau_i^* = (\tau_i, t_{i+1})$, in which $\tau_i = (j_1h, ..., j_ih)$ is fixed. Suppose t_{i+1} is a variable taking value on a closed interval $[a, b]$. We shall say the magnitude order is always the same if all elements $j_kh \notin (a, b)$, $k = 1, 2, ..., i$. In other words, if there is an element j_kh, such that both $t_{i+1} < j_kh$ and $j_kh < t_{i+1}$ can happen, then the magnitude order can change. In our problem, we have $j_kh \in \{0, h, 2h, ..., T\}$ fixed, and $t_{i+1} \in [(l-1)h, lh]$. It is clear that the analytical formula of the representation would not change when t_{i+1} varies. It ensures the representation, the right side of (3.11), is a twice continuously differentiable function of t_{i+1} on $[(l-1)h, lh]$, and thus we have local smoothness property $I_{\tau_i}(t_{i+1}) \in C^2[(l-1)h, lh]$ satisfied. However, when we consider the higher order quadrature rule, such as the Simpson's rule, the analytical formula of its representation can change. This is the reason why the extension of the analysis for the general quadrature rule is difficult.

We generate the lemma and theorems below based on our discussion.

Lemma 3.2.2. *Under assumption 3.2, for all $l = 1, 2, ..., T/h$, all $i = 0, 1, ..., m-1$ and all τ_i given, we have*

$$|G_{l,\tau_i} - Q_{l,\tau_i}| \leq L_i h^3,$$

where L_i is a constant with i fixed.

Proof. It is well-known that if the function $g : [t, t+h] \to \mathrm{R}$ is twice continuously differentiable, then we have

$$\left| \int_t^{t+h} g(x) \mathrm{d}x - \frac{h}{2} \left(g(t) + g(t+h) \right) \right| \leq \frac{h^3}{12} \sup_{x \in [t, t+h]} |g^{''}(x)|.$$

Since $I_{\tau_i}(t_{i+1})$ is a function of t_{i+1}, G_{l,τ_i} is an integral of $I_{\tau_i}(t_{i+1})$ over $t_{i+1} \in [(l-1)h, lh]$, and Q_{l,τ_i} is a trapezoidal approximation on G_{l,τ_i} with step size h, we can have a similar error bound provided that

$$I_{\tau_i}(t_i) \in C^2[(l-1)h, lh], \tag{3.12}$$

for any l, i and τ_i given. When $i = m-1$, we get

$$I_{\tau_{m-1}}(t_m) = f_m(t_{j_1}, t_{j_2}, ..., t_{j_m}),$$

where $t_{j_1}, t_{j_2}, ..., t_{j_m}$ is the re-order of t_m and all components in τ_{m-1}, such that $t_{j_\alpha} \leq t_{j_\beta}$ if $\alpha < \beta$. Thus, we have $I_{\tau_{m-1}}(t_m) \in C^2[(l-1)h, lh]$. When $i = 0, 1, ..., m-2$, we have discussed that $I_{\tau_i}(t_{i+1})$ can be written as

$$\sum_{\{D_{\tau_i^*}\}} \int_{D_{\tau_i^*}} f_m,$$

which is a twice continuously differentiable function of $t_{i+1} \in [(l-1)h, lh]$, as $\frac{\partial^2 f_m}{\partial t_j^2}$ is continuous for any j over the domain Ω_m. Then, the smoothness property (3.12) is justified, and it follows that

$$|G_{l,\tau_i} - Q_{l,\tau_i}| \leq \frac{h^3}{12} \sup_{t_{i+1} \in [(l-1)h, lh]} |I_{\tau_i}^{''}(t_{i+1})|, \tag{3.13}$$

for any l, i and τ_i.

Next, we have to find a constant E_i such that

$$\sup_{t_{i+1}\in[(l-1)h,lh]} |I''_{\tau_i}(t_{i+1})| \leq E_i,$$

for all l and τ_i. For $i = m-1$, we have

$$E_{m-1} \triangleq \sup_{\tau^*_{m-1}\in[0,T]^m} \left| \frac{\partial^2}{\partial t_m^2} f_m \right|,$$

and for $i = 0, 1, ..., m-2$, we let

$$E_i \triangleq \sup_{\tau^*_i\in[0,T]^{(i+1)}} \left| \frac{\partial^2}{\partial t_{i+1}^2} \sum_{\{D_{\tau^*_i}\}} \int_{D_{\tau^*_i}} f_m \right|,$$

where each element jh and t_{i+1} inside $\tau^*_i = (\tau_i, t_{i+1})$ are regarded as continuous variables on $[0,T]$, when we calculate the supremum norm. The supremum exists and is finite because $\frac{\partial^2}{\partial t_m^2} f_m$ and $\frac{\partial^2}{\partial t_{i+1}^2} \int_{D_{\tau^*_i}} f_m$ are continuous on the closed domain Ω_m. Then, $\sup_{t_{i+1}\in[(l-1)h,lh]} |I''_{\tau_i}(t_{i+1})|$ is bounded by E_i, and by (3.13), we obtain

$$|G_{l,\tau_i} - Q_{l,\tau_i}| \leq \frac{h^3}{12} E_i,$$

for all l and τ_i. We let $L_i = \frac{1}{12} E_i$, and the proof is complete. □

Theorem 3.2.1. *Let $\mathcal{T}\bar{f}_m$ be the trapezoidal approximation on the integral $\int_{\bar{\Omega}_m} \bar{f}_m$ with step size h. Under assumption 3.2, we have*

$$\left| \int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m \right| = O(h^2).$$

Proof. As discussed, we have the decomposition (3.8) for $\left| \int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m \right|$. It can be observed that these items in $|.|$ of (3.8) can be written as

$$\left| \sum_{l=1}^{T/h} (G_{l,\tau_i} - Q_{l,\tau_i}) \right| \leqslant T L_i h^2,$$

where the inequality is due to Lemma 3.2.2. It follows by (3.9) that $\left|\int_{\bar{\Omega}_m} \bar{f}_m - \mathcal{T}\bar{f}_m\right|$ has the error bound

$$\sum_{i=0}^{m-1} T^{i+1} L_i h^2 = O(h^2),$$

which completes the proof. □

Theorem 3.2.2. *Under assumptions 3.2 and 3.3, we have*

$$\mathbb{E}\left[\varphi_T \left(\fint_0^T \phi_t^h \mathrm{d}t\right)^m\right] = \mathbb{E}\left[\varphi_T \left(\int_0^T \phi_t \mathrm{d}t\right)^m\right] + O(h^2).$$

Proof. We can write

$$\mathbb{E}\left[\varphi_T \left(\int_0^T \phi_t \mathrm{d}t\right)^m\right] = \int_0^T ... \int_0^T \mathbb{E}(\phi_{t_1}...\phi_{t_m}\varphi_T)\mathrm{d}t_1...\mathrm{d}t_m \triangleq \int_{\bar{\Omega}_m} \bar{f}_m,$$

where the exchange of the expectation and the integration is justified by the Fubini theorem. Further, we denote $\mathcal{T}\bar{f}_m$ by the trapezoidal approximation on $\int_{\bar{\Omega}_m} \bar{f}_m$ with step size h, and it follows

$$\mathbb{E}\left[\varphi_T \left(\fint_0^T \phi_t^h \mathrm{d}t\right)^m\right] = \mathcal{T}\bar{f}_m.$$

The application of Theorem 3.2.1 finishes the proof. □

3.3 Applications

In this section, we apply the theorem of the stochastic trapezoidal discretization we have established for SDEs, which include the Heston model and the Stochastic Volatility with Jumps (SVJ) model.

3.3.1 Application on Heston Model

For the Heston model, recall that the model dynamics is as follows under the risk-neutral measure

$$\mathrm{d}S_t = rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) \tag{3.14}$$

$$\mathrm{d}V_t = k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1, \tag{3.15}$$

where $(S_t)_{t\geq 0}$ and $(V_t)_{t\geq 0}$ are the asset process and the variance process respectively. The parameters r, k, θ, σ an ρ are constants, and $(W_t^1)_{t\geq 0}$ and $(W_t^2)_{t\geq 0}$ are two independent Brownian motions.

We let $Y_t = \ln(S_t e^{-rt})$, and with the aid of the Itô formula, it follows by (3.14) that

$$Y_{t+h} = Y_t - \frac{1}{2}\int_t^{t+h} V_s\mathrm{d}s + \rho\int_t^{t+h}\sqrt{V_s}\mathrm{d}W_s^1 + \sqrt{1-\rho^2}\int_t^{t+h}\sqrt{V_s}\mathrm{d}W_s^2. \tag{3.16}$$

From (3.15), we have

$$\int_t^{t+h}\sqrt{V_s}\mathrm{d}W_s^1 = \frac{1}{\sigma}\left(V_{t+h} - V_t - k\theta h + k\int_t^{t+h} V_s\mathrm{d}s\right). \tag{3.17}$$

As the processes V and W^2 are mutually independent, $\int_t^{t+h}\sqrt{V_s}\mathrm{d}W_s^2$ in (3.16) is Normally distributed with the variance $\int_t^{t+h} V_s\mathrm{d}s$, given the path of V. Substituting (3.17) into (3.16), we obtain

$$Y_{t+h} = Y_t + \frac{\rho}{\sigma}(V_{t+h} - V_t - k\theta h) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_t^{t+h} V_s\mathrm{d}s + \sqrt{1-\rho^2}\sqrt{\int_t^{t+h} V_s\mathrm{d}s}N, \tag{3.18}$$

where N is a standard Normal random variable, independent of the variance process V. The formula (3.18) is due to Broadie & Kaya (2006). We employ the stochastic trapezoidal approximation

$$\int_t^{t+h} V_s\mathrm{d}s \approx \frac{V_t + V_{t+h}}{2}h \tag{3.19}$$

for equation (3.18), and we denote $\hat{Y}_{T,h}$ by the stochastic trapezoidal approximation of Y_T at time T with step size h. Then, we have the following lemma (Dufresne 2001, Theorem 2.3) saying the moment of the variance process is a smooth function of time. A function is called smooth if it is infinitely differentiable. Based on this result, we derive the weak convergence rate for the Heston model.

Lemma 3.3.1. *For the process $(V_t)_{t\geq 0}$ satisfying (3.15) with the initial V_0 given, the moment is*

$$\mathbb{E}(V_t^m|V_0) = \sum_{j=0}^{m} V_0^j g_{j,m}(t),$$

where $g_{j,m}(t)$ is a smooth function of t, over $t \in [0, +\infty)$, for any j and m.

Theorem 3.3.1. *Let $p(.)$ be any polynomial function, and then, we have*

$$\mathbb{E}(p(\hat{Y}_{T,h})) - \mathbb{E}(p(Y_T)) = O(h^2),$$

where h is the time step size.

Proof. Let N_i, $i = 1, ..., T/h$ be a series of mutually independent standard Normal random variables, that are also independent of the variance process V. Then, for any path of V given, there exists a random variable Z with the standard Normal distribution, such that

$$\sum_{i=1}^{T/h}\left(\sqrt{\int_{(i-1)h}^{ih} V_s \mathrm{d}s N_i}\right) = \sqrt{\int_0^T V_s \mathrm{d}s Z}.$$

Note that for each path of V, the random variable Z has the same distribution, and thus Z is independent of V. From (3.18), we have

$$Y_T = Y_0 + \frac{\rho}{\sigma}(V_T - V_0 - k\theta T) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_0^T V_s \mathrm{d}s + \sqrt{1-\rho^2}\sum_{i=1}^{T/h}\left(\sqrt{\int_{(i-1)h}^{ih} V_s \mathrm{d}s N_i}\right).$$

With substitution, we obtain

$$Y_T = Y_0 + \frac{\rho}{\sigma}(V_T - V_0 - k\theta T) + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_0^T V_s \mathrm{d}s + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s \mathrm{d}s Z}, \quad (3.20)$$

and the expression for $\hat{Y}_{T,h}$ is similar. Since any odd order moment of Z is zero, we have

$$\mathbb{E}\left[\left(\sqrt{\int_0^T V_s \mathrm{d}s Z}\right)^l\right] = \mathbb{E}\left[\left(\sqrt{\int_0^T V_s \mathrm{d}s}\right)^l\right]\mathbb{E}(Z^l) = 0,$$

where l is an odd number. We use the binomial expansion on $p(Y_T)$ and $p(\hat{Y}_{T,h})$, and it follows by (3.20) that $\mathbb{E}(p(\hat{Y}_{T,h})) - \mathbb{E}(p(Y_T))$ can be expressed as the summation of a finite number of terms (3.21) below with different combinations of non-negative integers m and n, up to scale factors

$$\mathbb{E}\left[V_T^n\left(\int_0^T V_t \mathrm{d}t\right)^m\right] - \mathbb{E}\left[V_T^n\left(⨍_0^T V_t^h \mathrm{d}t\right)^m\right], \tag{3.21}$$

where $⨍_0^T V_t^h \mathrm{d}t$ is the trapezoidal approximation on $\int_0^T V_t \mathrm{d}t$ with step size h.

We apply the theorem of stochastic trapezoidal discretization to analyse (3.21). By Lemma 3.3.1 and the tower rule, we have

$$\begin{aligned}
&\mathbb{E}(V_{t_1}...V_{t_m}V_T^n)\\
&= \mathbb{E}(\mathbb{E}(V_T^n|V_{t_m})V_{t_1}...V_{t_m})\\
&= \mathbb{E}\left(\sum_{j_m=0}^{n} g_{j_m,n}(T-t_m)V_{t_m}^{j_m+1}V_{t_1}...V_{t_{m-1}}\right)\\
&= \mathbb{E}\left(\sum_{j_m=0}^{n} g_{j_m,n}(T-t_m)\mathbb{E}(V_{t_m}^{j_m+1}|V_{t_{m-1}})V_{t_1}...V_{t_{m-1}}\right)\\
&= \sum_{j_m=0}^{n}\sum_{j_{m-1}=0}^{j_m+1}...\sum_{j_0=0}^{j_1+1} g_{j_m,n}(T-t_m)g_{j_{m-1},j_m+1}(t_m-t_{m-1})...g_{j_1,j_2+1}(t_2-t_1)g_{j_0,j_1+1}(t_1)V_0^{j_0},
\end{aligned} \tag{3.22}$$

where $0 \le t_1 \le ... \le t_m \le T$, and $g_{j,i}(t)$ is a smooth function of $t \in [0,\infty)$. Thus

$$\mathbb{E}(V_{t_1}...V_{t_m}V_T^n)$$

is also smooth over the domain $\Omega_m = \{(t_1,...,t_m)|0 \le t_1 \le ... \le t_m \le T\}$. Further, the variance process is always non-negative as discussed, so $\mathbb{E}|V_{t_1}...V_{t_m}V_T^n| = \mathbb{E}(V_{t_1}...V_{t_m}V_T^n)$. As $\mathbb{E}|V_{t_1}...V_{t_m}V_T^n|$ is a continuous function on the domain Ω_m, the condition

$$\int_{\Omega_m} \mathbb{E}|V_{t_1}...V_{t_m}V_T^n| < \infty$$

is also satisfied. Therefore, by Theorem 3.2.2, for any non-negative integers m and n, all of the terms (3.21) are $O(h^2)$, so does their finite summation. Hence, the proof is complete. □

By using the extrapolation, it is possible to eliminate the leading error term and obtain a higher weak order scheme, but it depends on the mathematical structure of the approximation error of the original scheme. Talay & Tubaro (1990) showed that for a weak-order-one numerical scheme, one can apply the extrapolation method to improve the weak order to two, provided the coefficients of the underlying SDE satisfy some classical conditions. The result was extended by Kloeden, Platen & Hofmann (1995) for higher-order weak extrapolation methods. In our analysis for the weak convergence, the error structure is similar to that of the trapezoidal rule on a multiple integral of a smooth integrand over a cube.

On the other hand, in the Heston model, higher moments of the asset price S_T might be infinite, see Andersen and Piterbarg (2006). The analysis of their convergence is not always meaningful, as the option price is usually well-defined and finite. Nevertheless, by analysing the structure of the approximation error more specifically and by using the Taylor expansion

$$S_T = \sum_{i=0}^{\infty} \frac{1}{i!} (\ln S_T)^i,$$

we may be able to extend the convergence result for S_T. We leave it for the future research.

3.3.2 Application on SVJ Model

There is an extension of the Heston model by adding a jump into the asset process $(S_t)_{t\geq 0}$, usually called the Stochastic Volatility with Jumps (SVJ) Model. Sometimes, it is referred as the Bates model, as it was introduced by Bates (1996) to deal with options of deutsche mark. 'The jump component can capture event-driven uncertainties, such as corporate defaults, operational failures or insured event (Platen & Bruti-Liberati 2010, Preface).' The SDE of the SVJ model under the risk-neutral measure is as follow

$$
\begin{aligned}
\mathrm{d}S_t &= (r - \lambda\bar{\mu})S_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) + (e^{J_{N(t)}} - 1)S_{t-}\mathrm{dN(t)} \\
\mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1,
\end{aligned}
$$

where λ is the constant intensity of the Poisson process $(N(t))_{t\geq 0}$. The random variable J_i denotes the i-th jump size, and it is Normally distributed with constant mean μ_J and variance σ_J^2 . The parameter $\bar{\mu}$ is specified to ensure $e^{-rt}S_t$ is a martingale, and it satisfies $\ln(1+\bar{\mu}) = \mu_J + \frac{1}{2}\sigma_J^2$.

Let $X_t = \ln(e^{-rt}S_t)$, and it follows by Itô formula for jump processes that

$$
\mathrm{d}X_t = -\left(\frac{1}{2} + \lambda\bar{\mu}\right)\mathrm{V_t}\mathrm{d}t + \rho\sqrt{V_t}\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\sqrt{V_t}\mathrm{d}W_t^2 + J_{N(t)}dN(t). \tag{3.23}
$$

To sample $(X_t)_{t\geq 0}$, the jump part and the diffusion part can be simulated separately, as both $(J_{N(t)})_{t\geq 0}$ and $(N(t))_{t\geq 0}$ in the jump of (3.23) are independent of the other elements of the SDE. In other words, one can first simulate the diffusion part of $(X_t)_{t\geq 0}$, and then add the jump sizes. The diffusion part can be simulated with any method for the Heston model we have discussed. The jump part can be simulated exactly, as the simulation only involves the Normal random variable J_i and the Poisson process $(N(t))_{t\geq 0}$. Similar to the discretization (3.18) as for the Heston model, we have

$$
\begin{aligned}
X_{t+h} = X_t &+ \frac{\rho}{\sigma}(V_{t+h} - V_t - k\theta h) + \left(\frac{\rho k}{\sigma} - \frac{1}{2} - \lambda\bar{\mu}\right)\int_t^{t+h} V_s\mathrm{d}s \\
&+ \sqrt{1-\rho^2}\sqrt{\int_t^{t+h} V_s\mathrm{d}s}N + \sum_{i=N(t^+)}^{N(t+h)} J_i - J_{N(t^+)},
\end{aligned} \tag{3.24}
$$

where we let $J_0 = 0$, and $N(t^+) \triangleq \lim_{l\to t^+} N(l)$. Here, $\lim_{l\to t^+}$ is for the right-sided limit, the limit when l decreases in value to approach t.

Denote $\hat{X}_{T,h}$ by the stochastic trapezoidal approximation on X_T at time T with step size h, and we obtain the weak convergence result below.

Theorem 3.3.2. *Let $p(.)$ be any polynomial function, and then, we have*

$$\mathbb{E}(p(\hat{X}_{T,h})) - \mathbb{E}(p(X_T)) = O(h^2),$$

where h is the time step size.

Proof. As the jump part in (3.24) only relies on the Poisson process $(N(t))_{t\geq 0}$ and the jump size J_i, that are independent of the other elements in the SDE, the proof is analogous to that of Theorem 3.3.1. To be more specific, we can write $X_T = Y_T + L$ and $\hat{X}_{T,h} = \hat{Y}_{T,h} + L$, where Y_T and $\hat{Y}_{T,h}$ are the Heston solutions we have discussed in the previous section. Here, $L = \sum_{i=N(t^+)}^{N(t+h)} J_i - J_{N(t^+)}$. Then, it follows that

$$\mathbb{E}(X_T^l) - \mathbb{E}(\hat{X}_{T,h}^l) = \sum_{i=0}^{l} \binom{l}{i} \mathbb{E}(L^{l-i})(\mathbb{E}(Y_T^i) - \mathbb{E}(\hat{Y}_{T,h}^i)),$$

for any positive integer l, and the application of Theorem 3.3.1 completes the proof. $\square$

Chapter 4

Multi-level Monte Carlo for the Heston Model

4.1 Introduction

The strong convergence result is important, in particular when a numerical scheme is implemented with the Multi-level Monte Carlo method. In this chapter, we shall introduce the Multi-level Monte Carlo (MLMC) version of the stochastic trapezoidal discretization for the Heston model, conditioned on that the variance process is simulated exactly. We consider different numerical schemes in the case of path-independent simulation, and in that of path-dependent simulation. In both situations, the MLMC estimators are defined, and the analytical convergence rates are derived within the full parameter regime, which keep in line with the numerical rates. Specifically, the theoretical convergence rate of the variance of the MLMC estimator we define is two for the path-independent simulation. For the path-dependent simulation, the rate is one under some restriction on the parameters, and it is half for all parameter regimes. These convergence rates are essential to estimate the computational complexity of MLMC.

In the literature, there are some discussions about the MLMC on Euler type discretizations for the Heston model, for instance, Giles (2008*b*), Kloeden & Neuenkirch (2012) and Altmayer & Neuenkirch (2015), and on Milstein type discretizations, for example, Giles & Szpruch (2014). However, their analysis is restricted to the parameter regime where the variance process can never hit zero. In particular, in Altmayer & Neuenkirch (2015), the authors considered the MLMC on the multi-dimensional Heston model for options with discontinuous payoffs. Further, we notice in these works, the original schemes, which the MLMC is based on, are constrained by the model parameters. Thus, it is reasonable that their MLMC versions would also have this limitation. To avoid this problem, we consider to establish the MLMC on a numerical scheme that applies without any restriction on the model parameters.

4.2 Multi-level Monte Carlo

The Multi-level Monte Carlo (MLMC) can be regarded as a variance reduction technique of the standard Monte Carlo, introduced by Giles (2008*b*). In this subsection, we shall look into it in the perspective of SDEs, although the technique can be used in a bigger context, such as SPDE and finite difference. Let P_T denote the payoff of an option at the maturity T, or sometimes just P for notational convenience, and $\hat{P}_{T,h}$ denote its approximation at time T with the step size h. In the standard Monte Carlo, we have

$$\mathbb{E}(P_T) \approx \frac{1}{n}\sum_{i=1}^{n} \hat{P}^i_{T,h},$$

where the number of sample n and the size of time step h are specified by the user. The mean square error, as we have discussed, is

$$\text{MSE}(\hat{e}) = O(n^{-1}) + O(h^{2\beta})$$

under some regular conditions on the drift and diffusion coefficients, where β is the corresponding weak convergence rate. The regular conditions include, but are not limited to, the Lipschitz condition and the linear growth condition. According to Duffie & Glynn (1995), the optimal tradeoff between h and n should be $n = O(h^{-2\beta})$. This means for a Euler scheme with the standard weak order one, in order to make the mean square error $O(\xi^2)$, one has to choose $n = O(\xi^{-2})$ and $h = O(\xi)$. So, the computational complexity is $O(\xi^{-3})$. However, the computational complexity of such a Euler scheme can be reduced to $O(\xi^{-2}(\ln \xi)^2)$ through the use of Multi-level

Monte Carlo, as was first discussed in Giles (2008*b*). The Multi-level Monte Carlo uses a linear combination of approximations with a series of times steps, rather than employs an approximation with a single time step as in the standard Monte Carlo. More samples are simulated on coarse levels of larger step sizes with low accuracy but less computational cost, whereas less samples are drawn on fine levels of smaller step sizes with high accuracy but more cost. In this way, it minimizes the overall computational cost with a given accuracy.

To see how it works, let $\hat{P}_l$ denote the approximation of P on level l, which corresponds to the step size $h_l = M^{-l}T$ with M a constant. The expected value $\mathbb{E}(\hat{P}_L)$ can be written as

$$\mathbb{E}(\hat{P}_L) = \mathbb{E}(\hat{P}_0) + \sum_{l=1}^{L} \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}).$$

MLMC estimates each of the expectations independently, but we shall notice within the same expectation $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$, the computation of $\hat{P}_l$ and $\hat{P}_{l-1}$ is not independent. Let Y_0 be an estimator for $\mathbb{E}(\hat{P}_0)$ using N_0 samples, and Y_l be an estimator for $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ using N_l paths, $l > 0$. One can define Y_l simply as

$$Y_l = N_l^{-1} \sum_{i=1}^{N_l} (\hat{P}_l^{(i)} - \hat{P}_{l-1}^{(i)}), \tag{4.1}$$

where N_l is the number of samples corresponding to the level l. Then, the combined MLMC estimator Y for $\mathbb{E}(\hat{P}_L)$ is

$$Y = \sum_{l=0}^{L} Y_l. \tag{4.2}$$

The estimated variance of Y_l, $l \geq 0$, is $\mathrm{Var}(Y_l) = N_l^{-1} V_l$, where $V_l = \mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$. So, we have $\mathrm{Var}(Y) = \sum_{l=0}^{L} N_l^{-1} V_l$. The computational complexity, that is to quantify the amount of time used for the MLMC algorithm to get the option price, can be defined as

$$N_0 + \sum_{l=1}^{L} N_l (M^l + M^{l-1}),$$

if one ignores the asymptotically negligible cost of the final payoff evaluation. Since

$h_l = M^{-l}T$, the computational complexity is proportional to

$$\sum_{l=0}^{L} N_l h_l^{-1}. \tag{4.3}$$

The analysis is generalized in the following theorem, see Giles (2008*b*), based on the computational complexity formula (4.3):

Theorem 4.2.1. (Giles) *Let P denote a functional of the solution of stochastic differential equation for a given Brownian path, and $\hat{P}_l$ denote the corresponding approximation using a numerical discretisation with the time step $h_l = M^{-l}T$.*

If there exist independent estimators Y_l based on N_l Monte Carlo samples, and positive constant $\beta \geq \frac{1}{2}$, α, c_1, c_2, c_3 such that

(i) $\mathbb{E}(\hat{P}_l - P) \leq c_1 h_l^{\beta}$

(ii) $\mathbb{E}(Y_l) = \begin{cases} \mathbb{E}(\hat{P}_0), & l = 0 \\ \mathbb{E}(\hat{P}_l - \hat{P}_{l-1}), & l > 0 \end{cases}$

(iii) $\mathrm{Var}(Y_l) \leq c_2 N_l^{-1} h_l^{\alpha}$

(iv) C_l, the computational complexity of Y_l, is bounded by

$$C_l \leq c_3 N_l h_l^{-1},$$

then there exists a positive constant c_4 such that for any $\xi < e^{-1}$ there are values L and N_l for which the multilevel estimator (4.2) has a mean-square-error with the bound

$$\mathrm{MSE} = \mathbb{E}[(Y - \mathbb{E}(P))^2] < \xi^2$$

with a computational complexity C with bound

$$C \leq \begin{cases} c_4 \xi^{-2}, & \alpha > 1 \\ c_4 \xi^{-2} (\ln \xi)^2, & \alpha = 1 \\ c_4 \xi^{-2-(1-\alpha)/\beta}, & 0 < \alpha < 1. \end{cases}$$

Giles (2008*b*) further showed that the optimal N_l in (4.1) is determined by

$$N_l = \left\lceil 2\xi^{-2}\sqrt{V_l h_l}\left(\sum_{l=0}^{L}\sqrt{V_l/h_l}\right)\right\rceil, \tag{4.4}$$

so that $V(Y) < \frac{1}{2}\xi^2$, where $\lceil x \rceil$ represents the smallest integer that is no smaller than x. For the standard Monte Carlo on a time-discrete scheme, the simulation accuracy is determined by the step size and the number of paths, whereas for MLMC, the step size and number of paths are optimized in the MLMC algorithm, and the accuracy is only controlled by the input parameter ξ in the theorem above. Except for establishing the correct exponent β for condition (i), the challenge will be in determining and proving the appropriate exponent α for (iii).

The convergence rate α is highly important as it determines the computational complexity C, as we can see from the the theorem. The weak convergence rate β only influences the order of computational complexity when $0 < \alpha < 1$. To calculate the expectation $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$, we shall first simulate a sample $\hat{P}_l$ on a fine level and then compute $\hat{P}_{l-1}$ on a coarse level, based on the information we obtained when simulating $\hat{P}_l$. The standard MLMC estimator is to keep $\hat{P}_{l-1}$ and $\hat{P}_l$ sampled from the same Brownian motion path. However, there is a freedom as how to sample $\hat{P}_{l-1}$ conditioned on $\hat{P}_l$ sampled, and clearly different ways of sampling may lead to different values of α. We prefer a MLMC estimator with large α for small complexity C. To avoid the introduction of additional bias, we shall make sure that $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ and in $\mathbb{E}(\hat{P}_{l+1} - \hat{P}_l)$ have the same expectation $\mathbb{E}(\hat{P}_l)$, as was firstly discussed inGiles (2008*a*).

Since the breakthrough by Giles, the MLMC has found a huge application for many time-discrete schemes. For the Euler scheme, see Giles (2008*b*), and for the Milstein scheme, we refer to Giles (2008*a*). An overview of the progress on MLMC can be found in Giles & Szpruch (2013). MLMC is particularly useful in financial engineering, where there is a very high demand for computational efficiency. It was reported in Giles (2008*b*) that for a Euler scheme, it is possible to reduce the computational complexity by a factor 100 through the use of MLMC, compared to the standard Monte Carlo. This depends on the model parameters and the payoff function of an option. In terms of the MLMC analysis, the theoretical strong convergence rates of the Euler scheme for options with various payoff functions, including Lipschitz, Asian, Lookback, Barrier and Digital, can be found at Giles, Higham & Mao (2009) and these of the Milstein scheme can be seen at Giles, Debrabant & Rößler (2013). Avikainen (2009) improved the result of Giles et al. (2009) by providing a sharper error bound for the Digital option. However, these results are based on the assumption that the coefficients of SDEs are Lipschitz continuous, which is not satisfied with many SDEs in financial engineering. Further, MLMC can be generalized by starting with a level 0 of multiple steps. Let $h_0^{'} = TH^{-1}$ with H a positive integer, and let $h_l^{'} = M^{-l}TH^{-1}$. The theorem remains true when h_l is replaced by $h_l^{'}$, and the proof is straightforward following the proof of the original theorem of Giles. We shall discuss in the last chapter the application of this result when we implement MLMC for the Heston model with piece-wise constant parameters.

Finally, we will give an algorithm for the MLMC we have discussed, which is a trivial extension of the MLMC algorithm in Giles (2008*b*) for the Euler scheme. For a numerical scheme on SDEs with the weak convergence rate β, asymptotically, as $l \to \infty$, we have

$$\mathbb{E}(P - \hat{P}_l) \approx c_1 h_l^{\beta},$$

for some constant c_1, and hence

$$\begin{aligned}\mathbb{E}(\hat{P}_l - \hat{P}_{l-1}) &= \mathbb{E}(P - \hat{P}_{l-1}) - \mathbb{E}(P - \hat{P}_l)\\ &\approx (M^{\beta} - 1)c_1 h_l^{\beta}\\ &\approx (M^{\beta} - 1)E(P - P_l). \end{aligned} \tag{4.5}$$

One increases the value for L until

$$|Y_L| < \frac{1}{\sqrt{2}}(M^{\beta} - 1)\xi. \tag{4.6}$$

From (4.5), the magnitude of the bias is less than $\xi/\sqrt{2}$. The choice of optimal N_l in (4.4) guarantees $V(Y) < \frac{1}{2}\xi^2$. Therefore in a theoretical perspective, it should give a mean square error that is less than ξ^2. However, in practice, the variance V_l

required to compute N_l is estimated by a certain number of samples embedded in the algorithm, which may not be sufficiently large, and thus there is no guarantee to achieve a MSE error less than ξ^2. The numerical algorithm is provided as follows

1. Start with $L = 0$.
2. Estimate V_L using an initial set of $N_L = 1000$ samples.
3. Define optimal N_l, $l = 0, ..., L$, using equation (4.4).
4. Evaluate extra samples at each level as needed for new N_l.
5. $L \geq 2$, test for convergence using the formula (4.6).
6. If $L < 2$, or it is not converged, set $L = L + 1$ and go to Step 2.

4.3 Novel MLMC Estimators for the Heston Model

Recall that the dynamics of the Heston model is as follows

$$\begin{aligned}
\mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) \\
\mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1.
\end{aligned}$$

As discussed, we have such an exact formula

$$\begin{aligned}
S_t = S_u \exp\Bigg[& r(t-u) - \frac{1}{2}\int_u^t V_s\mathrm{d}s + \frac{\rho}{\sigma}\left(V_t - V_u - k\theta(t-u) + k\int_u^t V_s\mathrm{d}s\right) \\
& + \sqrt{1-\rho^2}\sqrt{\int_u^t V_s\mathrm{d}sN}\Bigg], \qquad (4.7)
\end{aligned}$$

where N represents a random variable with the standard Normal distribution, that is independent of the variance process. The integral $\int_u^t V_s\mathrm{d}s$ in (4.7) can be approximated by the stochastic trapezoidal rule. For any time $t \in [0, T]$, and any step size h, such that T/h is a positive integer, we simply denote

$$\fint_t^{t+h} V_s^h\mathrm{d}s \triangleq \frac{V_t + V_{t+h}}{2}h$$

and

$$\int_0^T V_s^h \mathrm{d}s \triangleq \sum_{i=1}^{T/h} \frac{V_{(i-1)h} + V_{ih}}{2} h.$$

We shall introduce the MLMC estimators, separated by the path-independent and the path-dependent simulations. As discussed in Chapter 1, the path-independent simulation is to price path-independent options, while the path-dependent simulation is for path-dependent options.

4.3.1 Simulation for Path-independent Options

Denote by $\tilde{S}_{T,h}$ the approximation of S_T at time T. For path-independent simulation, to calculate $\mathbb{E}(\tilde{S}_{T,Mh} - \tilde{S}_{T,h})$, we simply let the asset price $\tilde{S}_{T,Mh}$ at the coarse level keep the same standard Normal random variable N as required to simulate $\tilde{S}_{T,h}$ at the fine level.

Specifically, at the fine level, we have

$$\begin{aligned}\ln \tilde{S}_{T,h} = \ln S_0 + \Bigg[& rT - \frac{1}{2}\int_0^T V_s^h \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s^h \mathrm{d}s\right) \\ & + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s^h \mathrm{d}s N}\Bigg],\end{aligned}$$

while at the coarse level, we define

$$\begin{aligned}\ln \tilde{S}_{T,Mh} = \ln S_0 + \Bigg[& rT - \frac{1}{2}\int_0^T V_s^{Mh} \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s^{Mh} \mathrm{d}s\right) \\ & + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s^{Mh} \mathrm{d}s N}\Bigg],\end{aligned}$$

with the same N in both equations.

Remark 4.3.1. *We stress the fact that the two equations have the same N does not mean they use the same path of the Brownian motion $(W_t^2)_{t\geq 0}$. To better understand this, we can simplify the problem as follows: Consider an integral $\int_t^{t+h} \sqrt{V_s}\mathrm{d}W_s^2$, and we can write it as $\sqrt{\int_t^{t+h} V_s \mathrm{d}s}N$, with N a standard Normal random variable. Now, we approximate $\sqrt{\int_t^{t+h} V_s \mathrm{d}s}N$ as $\sqrt{\frac{V_t+V_{t+h}}{2}h}N$. Is it equivalent to approximating $\int_t^{t+h} \sqrt{V_s}\mathrm{d}W_s^2$ as $\int_t^{t+h} \sqrt{\frac{V_t+V_{t+h}}{2}}\mathrm{d}W_s^2$? At the first glance, the answer seems yes, as*

$\int_t^{t+h}\sqrt{\frac{V_t+V_{t+h}}{2}}\mathrm{d}W_s^2$ *has the mean 0 and the variance* $\frac{V_t+V_{t+h}}{2}h$*. By a careful analysis, we observe that in general the expectations*

$$\mathbb{E}\left[\left(\sqrt{\int_t^{t+h}V_s\mathrm{d}s}N-\sqrt{\frac{V_t+V_{t+h}}{2}h}N\right)^2\right]$$
$$\neq\mathbb{E}\left[\left(\int_t^{t+h}\sqrt{V_s}\mathrm{d}W_s^2-\int_t^{t+h}\sqrt{\frac{V_t+V_{t+h}}{2}}\mathrm{d}W_s^2\right)^2\right].$$

Thus, the answer is no. In other words, if we have to keep the same path of the Brownian motion $(W_t^2)_{t\geq 0}$*, we have to write* $\int_t^{t+h}\sqrt{\frac{V_t+V_{t+h}}{2}}\mathrm{d}W_s^2$ *as* $\sqrt{\frac{V_t+V_{t+h}}{2}h}Z$*, with* Z *a standard Normal random variable different from* N*. The link between* N *and* $(W_t^2)_{t\geq 0}$ *is not easy to describe, which we are not going to explore in this thesis. In the literature concerning MLMC, such as Giles (2008*b*) and Giles (2008*a*), the path of the Brownian motion is usually kept the same while one simulates at different levels. However, it is not a necessary condition as long as the variance* $\mathrm{Var}(\hat{P}_l-\hat{P}_{l-1})$ *converges. We believe this result would be useful for more applications of MLMC in the further research, if it is difficult to construct a MLMC estimator keeping the same Brownian motion path on the coarse level as on the fine level.*

One can verify that the payoff $\hat{P}_l$ when estimating $\mathbb{E}(\hat{P}_l-\hat{P}_{l-1})$ and $\mathbb{E}(\hat{P}_{l+1}-\hat{P}_l)$ have the same expectation, so no additional bias is implemented. On the other hand, the choice of M in the equation above has no influence on the convergence rate of $\mathrm{Var}(\hat{P}_l-\hat{P}_{l-1})$, which we shall show in the next subsection. According to the Theorem of Multi-level Monte Carlo we stated in the previous section, the choice of M will not affect the rate of computational complexity. However, it does affect the efficiency, the real computational time to get an option price of desired accuracy. Thus, one has to find out the optimal value of M through some numerical test.

Finally, we emphasize that the MLMC estimator we have defined applies to any payoff function of an option, as long as it is path-independent. However, in the theoretical analysis of its convergence rate, we have to impose the Lipschitz assumption on the payoff function.

4.3.2 Convergence Analysis for Path-independent Simulation

Let $P(.)$ be a Lipschitz payoff function, such that

$$|P(U) - P(V)| \leq c|U - V|$$

for all U, V, where c is a constant. Throughout the analysis, c represents a constant regardless its value, unless otherwise stated. Then, we have $\mathbb{E}\left[(P(U) - P(V))^2\right] \leq c\mathbb{E}\left[(U - V)^2\right]$, where $\mathbb{E}\left[(U - V)^2\right]$ is usually more convenient for the error analysis of the Lipschitz payoff. Under the Lipschitz assumption, for the payoffs of the majority of put options, we further have

$$\mathbb{E}\left[(P(U) - P(V))^2\right] \leq c\mathbb{E}\left[(\ln U - \ln V)^2\right].$$

In our case, we let $U = \tilde{S}_{T,Mh}$, and $V = \tilde{S}_{T,h}$. It follows that

$$\mathbb{E}\left[(P(\tilde{S}_{T,Mh}) - P(\tilde{S}_{T,h}))^2\right] \leq c\mathbb{E}\left[(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h})^2\right]. \tag{4.8}$$

We shall derive the convergence rate of the right side of this inequality, and then the convergence rate of

$$\mathrm{Var}(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}),$$

that is useful to estimate the computational complexity of the MLMC. For call options, the inequality (4.8) is usually not satisfied. However, for many types of options in finance, we have the put-call parity, which says the difference of prices between a call option and its corresponding put option with the same strike and the same maturity is equivalent to that of a single forward contact, known analytically. This means the computational complexity to calculate a call option price, through the use of the put-call parity, is just the same as that to compute its corresponding put option price.

Assumption Analysis

In this part, we shall provide more specific analysis on the assumption we impose, and have more discussion on the put-call parity. The theorem below says our assumption is typically for a put option with a Lipschitz payoff function.

Theorem 4.3.1. *Let $P : \mathbb{R}^+ \to \mathbb{R}^+$ be a Lipschitz continuous function. Suppose there is a constant L, and for all $U \geq L$, we have $P(U) = P(L)$. Then, for all U, V, there exists a constant C, such that*

$$|P(U) - P(V)| \leq C|\ln U - \ln V|.$$

Proof. We shall firstly consider the case where both $U, V \in [0, L]$, and it follows by the Lipschitz continuity that

$$|P(U) - P(V)| \leq c|U - V| \leq cL|\ln U - \ln V|. \tag{4.9}$$

Next, we consider the case where $U \in [0, L]$ and $V \in (L, \infty)$, and by (4.9), we have

$$|P(U) - P(V)| = |P(U) - P(L)| \leq cL|\ln U - \ln L| < cL|\ln U - \ln V|.$$

The analysis of the situation where $V \in [0, L]$ and $U \in (L, \infty)$ is analogous. When both $U, V \in (L, \infty)$, we obtain

$$|P(U) - P(V)| = 0.$$

Therefore, let $C = cL$, and we complete the proof. □

Remark 4.3.2. *In this theorem, we notice that the condition $P(U) = P(L)$, when $U \geq L$, is in particular relevant to the put option. The put option becomes worthless when the price of the underlying asset reaches a certain high level.*

The put-call parity is highly useful in practice. For example, we consider a call option with the payoff function $Pc(S_T) = (f(S_T) - K)^+$, where $f(.)$ can be any real function with no constant term (no cash flow) and K is the strike. Clearly, the corresponding put option has the payoff function $Pp(S_T) = (K - f(S_T))^+$. Denote by C_t and P_t the t time prices of the call and the put respectively, where $t \in [0, T]$. At time T, we have

$$C_T - P_T = f(S_T) - K.$$

According to the no-arbitrary assumption, which is a fundamental assumption in mathematical finance, this equation should also be true at any time $t \in [0, T]$, if there is no dividend payment. Then, we obtain

$$C_t - P_t = f(S_t) - e^{-r(T-t)}K.$$

This is exactly the put-call parity formula we obtain for our example. At time t, conditioned on S_t given, the right side of the equation is a value available at time t. Another example where the put-call parity exists is the European basket option. The payoff function of such a call option is of the form

$$\left(\sum_{i=1}^{n} \theta_i S_T^i - K\right)^+,$$

where $(\theta_1, ..., \theta_n)$ is the weight vector, with all of its elements in $[0, 1]$, and $(S_T^1, ..., S_T^n)$ is the multiple asset prices at maturity T. Denote by C_t^B the call option, and P_t^B by the put option at time t. Following the same analysis as in the previous example, we have

$$C_t^B - P_t^B = \sum_{i=1}^{n} \theta_i S_t^i - e^{-r(T-t)}K,$$

the right side of which is again observable at time t.

Let us go back to our problem: we aim to derive an error bound for such an expectation

$$\mathbb{E}\left[(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h})^2\right],$$

and we have the theorem below.

Theorem 4.3.2. *There is an inequality*

$$\mathbb{E}\left[\left(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}\right)^2\right] \leq E_I(h) + E_I(Mh),$$

where

$$E_I(\delta) = c\mathbb{E}\left[\left(\fint_0^T V_s^\delta \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2\right] + c\mathbb{E}\left[\left(\sqrt{\fint_0^T V_s^\delta \mathrm{d}s} - \sqrt{\int_0^T V_s \mathrm{d}s}\right)^2\right].$$

Proof. We have

$$\begin{aligned}
&\mathbb{E}\left[\left(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}\right)^2\right] \\
&\leq 2\mathbb{E}\left[\left(\ln \tilde{S}_{T,Mh} - \ln S_T\right)^2\right] + 2\mathbb{E}\left[\left(\ln \tilde{S}_{T,h} - \ln S_T\right)^2\right] \qquad (4.10)
\end{aligned}$$

where

$$\begin{aligned}
\ln S_T = \ln S_0 + &\left[rT - \frac{1}{2}\int_0^T V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s \mathrm{d}s\right)\right. \\
&\left. + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s \mathrm{d}s N}\right].
\end{aligned}$$

We use the same random variable N in $\tilde{S}_{T,h}$ and in $\tilde{S}_{T,Mh}$. Due to the independence of N and the variance process, it follows by the straight-forward calculation that

$$\begin{aligned}
\mathbb{E}\left[\left(\ln \tilde{S}_{T,h} - \ln S_T\right)^2\right] &= c\mathbb{E}\left[\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2\right] \\
&\quad + c\mathbb{E}\left[\left(\sqrt{\fint_0^T V_s^h \mathrm{d}s} - \sqrt{\int_0^T V_s \mathrm{d}s}\right)^2\right]\mathbb{E}(N^2). \qquad (4.11)
\end{aligned}$$

The analysis of $\mathbb{E}\left[(\ln \tilde{S}_{T,Mh} - \ln S_T)^2\right]$ is analogous by replacing h with Mh. The above inequality (4.11) together with (4.10) finishes the proof. □

Our analysis will be based on the inequality stated in the theorem above. Since in the path-independent simulation, the option price only depends on the asset process at maturity T, it suffices to analyse $\mathbb{E}\left[\left(\ln \tilde{S}_{T,Mh} - \ln \tilde{S}_{T,h}\right)^2\right]$ at time T.

The analysis relies on the structure of the variance process, which is the extension of Theorem 3.3.1 in the previous chapter. Specifically, it follows by (3.22) that

$$\mathbb{E}(V_{t_1}...V_{t_m}) = \sum_{j=0}^{m} p_{j,m}(t_1, t_2, ..., t_m) V_0^j, \tag{4.12}$$

where $p_{j,m}(.)$ is a smooth function of $t_1, t_2, ..., t_m$ satisfying $0 \leq t_1 \leq t_2 \leq ... \leq t_m \leq T$.

Theorem 4.3.3. *Let $(V_t)_{t\geq 0}$ be the variance process. Then, we have*

$$\mathbb{E}\left[\left(⨍_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2\right] = O(h^2).$$

Proof. There is an expansion

$$\begin{aligned}\mathbb{E}\left[\left(⨍_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^2\right] &= \mathbb{E}\left[\left(⨍_0^T V_s^h \mathrm{d}s\right)^2\right] - 2\mathbb{E}\left(⨍_0^T V_s^h \mathrm{d}s \cdot \int_0^T V_s \mathrm{d}s\right) \\ &\quad + \mathbb{E}\left[\left(\int_0^T V_s \mathrm{d}s\right)^2\right].\end{aligned} \tag{4.13}$$

It follows by Theorem 3.2.2 and the representation (4.12) that

$$\mathbb{E}\left[\left(⨍_0^T V_s^h \mathrm{d}s\right)^2\right] = \mathbb{E}\left[\left(\int_0^T V_s \mathrm{d}s\right)^2\right] + O(h^2). \tag{4.14}$$

Now it remains to consider $\mathbb{E}\left(⨍_0^T V_s^h \mathrm{d}s \cdot \int_0^T V_s \mathrm{d}s\right)$. It is well-known that, for a twice continuously differentiable function $f : [0, T] \to \mathbb{R}$, we have

$$\int_0^T f(x)\mathrm{d}x = ⨍_0^T f^h(x)dx + O(h^2).$$

Here, we let $f(x) = \mathbb{E}(V_x \int_0^T V_s \mathrm{d}s)$, and we have a decomposition

$$\mathbb{E}\left(V_x \int_0^T V_s \mathrm{d}s\right) = \int_0^x \mathbb{E}(V_x V_s)\mathrm{d}s + \int_x^T \mathbb{E}(V_x V_s)\mathrm{d}s,$$

where the exchange of the expectation and the integral is due to the Fubini theorem. The expectation $\mathbb{E}(V_x V_s)$ is a smooth function on $0 \leq s \leq x \leq T$ and it is also smooth on $0 \le x \le s \le T$, due to (4.12). The former ensures that the integral $\int_0^x \mathbb{E}(V_x V_s)\mathrm{d}s$ is a smooth function of $x \in [0,T]$, and the latter ensures the smoothness of $\int_x^T \mathbb{E}(V_x V_s)\mathrm{d}s$ on $x \in [0,T]$. Thus, it follows that $\mathbb{E}(V_x \int_0^T V_s \mathrm{d}s)$ is a smooth function of $x \in [0,T]$, and we have

$$
\begin{aligned}
\mathbb{E}\left(⨍_0^T V_s^h \mathrm{d}s \cdot \int_0^T V_s \mathrm{d}s\right) &= ⨍_0^T \mathbb{E}\left(V_x^h \int_0^T V_s \mathrm{d}s\right)\mathrm{d}x \\
&= \int_0^T \mathbb{E}\left(V_x \int_0^T V_s \mathrm{d}s\right)\mathrm{d}x + O(h^2) \\
&= \mathbb{E}\left[\left(\int_0^T V_s \mathrm{d}s\right)^2\right] + O(h^2), \qquad (4.15)
\end{aligned}
$$

where the exchange of the expectation and the integral is valid again due to the Fubini theorem. Equation (4.13) together with (4.14) and (4.15) finishes the proof.

□

Next, we try to extend the result by proving

$$
\mathbb{E}\left[\left(⨍_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n}\right] = O(h^{2n}),
$$

for any positive integer n. This requires a different approach. To prove it, we shall develop a theorem that applies in a much more general context.

Let $(X_t)_{t\ge 0}$ be a time-homogeneous stochastic process on probability space $(\Omega, \mathcal{F}, P)$ with the filtration $(\mathcal{F}_t)_{t\ge 0}$ satisfying the usual conditions. We consider a quadrature rule

$$
\int_t^{t+h} X_s \mathrm{d}s \approx (\eta X_t + (1-\eta)X_{t+h})h
$$

for some $\eta \in [0,1]$, where $t \in [0,T]$. If $\eta = 1$, it is an Euler rule, and if $\eta = 1/2$, it is a trapezoidal rule. For notational simplicity, we denote

$$
X_{t,h}(s) \triangleq X_s - (\eta X_t + (1-\eta)X_{t+h}),
$$

where $s \in [t, t+h]$. We can also define

$$
X_{t,h}(s_1, .., s_m) \triangleq X_{t,h}(s_1) \cdot \ldots \cdot X_{t,h}(s_m).
$$

Our theorem is based on the two assumptions below.

Assumption 4.3.1. *The expectation of the product of $(X_t)_{t\geq 0}$ at different time has the following form*

$$\mathbb{E}(X_{t_1}...X_{t_m}) = \sum_{j=0}^{m} p_{j,m}(t_1, t_2, ..., t_m) X_0^j, \tag{4.16}$$

where $p_{j,m}(.)$ is a continuously differentiable function of $t_1, t_2, ..., t_m$, in which we assume $0 \leq t_1 \leq t_2 \leq ... \leq t_m \leq T$.

Assumption 4.3.2. *Conditioned on X_t, we have*

$$\lim_{h\to 0} \mathbb{E}\left[|X_{t,h}(s)|^n | X_t\right] = 0$$

for any non-negative integer n and $s \in [t, t+h]$.

We shall prove that

$$\mathbb{E}\left[\left(\sum_{i=1}^{T/h}(\eta X_{(i-1)h} + (1-\eta)X_{ih})h - \int_0^T X_s \mathrm{d}s\right)^{2n}\right] = O(h^{2n}),$$

under these two assumptions. Typically, assumption 4.3.1 is for the structure of the expectation function and assumption 4.3.2 is for the convergence in a local sense.

Lemma 4.3.1. *Suppose X_t is given, and assumptions 4.3.1 and 4.3.2 are satisfied. Then, for any $s_1, ..., s_m \in [t, t+h]$, and any non-negative integer n, we have*

$$\mathbb{E}\left(X_{t+h}^n X_{t,h}(s_1) \cdot ... \cdot X_{t,h}(s_m) | X_t\right) = \sum_{j=0}^{m+n} L_h^1 X_t^j,$$

where L_h^1 is independent of X_t, and there exists a constant L, only depending on m and n, such that $|L_h^1| \leq Lh$ for all h.

Proof. Let $t_i \in \{t, s_i, t+h\}$ for any $i = 1, 2, ..., m$, and let $t_i = t+h$ for any $i = m+1, ..., m+n$. So, we have $t_i \in [t, t+h]$. Since $X_{t,h}(s) \triangleq X_s - (\eta X_t + (1-\eta)X_{t+h})$

for some $\eta \in [0,1]$, there is an expansion

$$\mathbb{E}\left(X^n_{t+h}X_{t,h}(s_1,..,s_m)|X_t\right) = \sum_{t_1,...,t_m} c_{t_1,...,t_m}\mathbb{E}(X_{t_1}...X_{t_{m+n}}|X_t), \tag{4.17}$$

where $c_{t_1,...,t_m}$ is a constant, and $|c_{t_1,...,t_m}| \le 1$ due to that $\eta \in [0,1]$. To analyse the expectation $\mathbb{E}(X_{t_1}...X_{t_{m+n}}|X_t)$, we assume without the loss of generality that $t_1 \le t_2 \le ... \le t_m$. Under assumption 4.3.1, we have

$$\begin{aligned}
&\mathbb{E}(X_{t_1}...X_{t_{m+n}}|X_t) \\
&= \sum_{j=0}^{m+n} p_{j,m+n}(t_1-t,...,t_{m+n}-t)X_t^j \\
&= \sum_{j=0}^{m+n}\left(p_{j,m+n}(0,...,0)+\sum_{i=1}^{m+n}\frac{\partial p_{j,m+n}(t_\xi)}{\partial t_i}(t_i-t)\right)X_t^j.
\end{aligned} \tag{4.18}$$

The last line is due to the Taylor expansion at the point $(0,...,0)$, and t_ξ is a $(m+n)$-dimensional vector on the line segment between $(0,...,0)$ and $(t_1-t,...,t_{m+n}-t)$. We combine (4.17) and (4.18) together, and obtain

$$\begin{aligned}
&\mathbb{E}\left(X^n_{t+h}X_{t,h}(s_1,..,s_m)|X_t\right) \\
&= \sum_{j=0}^{m+n}\sum_{t_1,...,t_m} c_{t_1,...,t_m}\left(p_{j,m+n}(0,...,0)+\sum_{i=1}^{m+n}\frac{\partial p_{j,m+n}(t_\xi)}{\partial t_i}(t_i-t)\right)X_t^j.
\end{aligned} \tag{4.19}$$

On the other hand, under assumption 4.3.2, it follows by the Generalized Hölder's inequality that

$$\begin{aligned}
&\lim_{h\to 0}\left|\mathbb{E}\left(X^n_{t+h}X_{t,h}(s_1,..,s_m)|X_t\right)\right| \\
&\le \lim_{h\to 0}\sqrt[m+n]{\mathbb{E}^n\left[|X_{t+h}|^{m+n}|X_t\right]\mathbb{E}\left[|X_{t,h}(s_1)|^{m+n}\middle| X_t\right]...\mathbb{E}\left[|X_{t,h}(s_m)|^{m+n}\middle| X_t\right]} = 0.
\end{aligned}$$

This suggests that the constant terms in (4.19) are always zero, and we can get

$$\sum_{t_1,...,t_m} c_{t_1,...,t_m}(p_{j,m+n}(0,...,0)) = 0$$

for any j. Therefore, we can simplify (4.19) as

$$\mathbb{E}\left(X_{t+h}^n X_{t,h}(s_1,..,s_m)|X_t\right) = \sum_{j=0}^{m+n}\left(\sum_{t_1,...,t_m} c_{t_1,...,t_m} \sum_{i=1}^{m+n} \frac{\partial p_{j,m+n}(t_\xi)}{\partial t_i}(t_i - t)\right) X_t^j. \tag{4.20}$$

Furthermore, as

$$\left|\frac{\partial p_{j,m+n}(t_\xi)}{\partial t_i}\right|$$

for any i and j is bounded in the simplex domain $\Omega_{m+n} = \{(l_1,...,l_{m+n})|0 \leq l_1 \leq ... \leq l_{m+n} \leq T\}$ due to the continuously differentiability of $p_{j,m+n}(.)$, and as $|t_i - t| \leq h$ and $|c_{t_1,...,t_m}| \leq 1$, we can write

$$\begin{aligned}
L_h^1 &\triangleq \sum_{t_1,...,t_m} c_{t_1,...,t_m}\left(\sum_{i=1}^{m+n}\frac{\partial p_{j,m+n}(t_\xi)}{\partial t_i}(t_i - t)\right) \\
&\leq \max_{j=0,1,...,m+n}\left(\sum_{t_1,...,t_m}|c_{t_1,...,t_m}|\sum_{i=1}^{m+n}\sup_{t_{i,j}\in\Omega_{m+n}}\left|\frac{\partial p_{j,m+n}(t_{i,j})}{\partial t_i}\right|\right) h \\
&< \max_{j=0,1,...,m+n}\left(3^m\sum_{i=1}^{m+n}\sup_{t_{i,j}\in\Omega_{m+n}}\left|\frac{\partial p_{j,m+n}(t_{i,j})}{\partial t_i}\right|\right) h \\
&\triangleq Lh.
\end{aligned}$$

We can see that L_h^1 is independent of X_t, and L does not rely on the choice of $s_1,...,s_m$, which only depends on m and n . With L_h^1 substituted into (4.20), the proof is complete. □

Lemma 4.3.2. *Suppose $t_i = a_i h$, for $i = 1,2,...,l$, where $a_i \in \{0,1,2,...,T/h-1\}$. We assume that $t_i \neq t_j$, if $i \neq j$. Then, under assumptions 4.3.1 and 4.3.2, we have*

$$\mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_l,h}(s_1^l,...,s_{m_l}^l)\right) = L_h^l,$$

where $s_1^i,...,s_{m_i}^i \in [t_i, t_i+h]$ for all $i = 1,...,l$, and there exists a constant L, solely relies on m_1,...,m_l such that $|L_h^l| \leq Lh^l$.

Proof. Without the loss of generalisation, we assume that $t_1 < t_2 < ... < t_l$. Under assumption 4.3.1, we know that the expectation

$$\mathbb{E}(X_t^j|X_0) = \sum_{i=0}^{j} p_{i,m}(t)X_0^i,$$

where $p_{i,m}(t)$ is a continuously differentiable function, and is thus bounded in $t \in [0,T]$. Based on our previous notation L_h^1, we define $L_h^2 \triangleq L_h^1 L_h^1$, $L_h^3 \triangleq L_h^2 L_h^1$,..., $L_h^l \triangleq L_h^{l-1} L_h^1$. Since in our consideration, $t_i \in \{0, h, 2h, ..., T\}$, by the Tower's rule and Lemma 4.3.1, we have

$$\begin{aligned}
&\mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_l,h}(s_1^l,...,s_{m_l}^l)\right) \\
&= \mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-1},h}(s_1^{l-1},...,s_{m_{l-1}}^{l-1})\mathbb{E}(X_{t_l,h}(s_1^l,...,s_{m_l}^l)|X_{t_l})\right) \\
&= \mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-1},h}(s_1^{l-1},...,s_{m_{l-1}}^{l-1})\sum_{j_l=0}^{m_l} L_h^1 X_{t_l}^{j_l}\right) \\
&= \mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-1},h}(s_1^{l-1},...,s_{m_{l-1}}^{l-1})\sum_{j_l=0}^{m_l} L_h^1 \mathbb{E}(X_{t_l}^{j_l}|X_{t_{l-1}+h})\right) \\
&= \mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-1},h}(s_1^{l-1},...,s_{m_{l-1}}^{l-1})\sum_{j_l=0}^{m_l}\sum_{i_l=0}^{j_l} L_h^1 X_{t_{l-1}+h}^{i_l}\right) \\
&= \mathbb{E}\Bigg(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-2},h}(s_1^{l-2},...,s_{m_{l-2}}^{l-2}) \\
&\quad \sum_{j_l=0}^{m_l}\sum_{i_l=0}^{j_l} L_h \mathbb{E}(X_{t_{l-1}+h}^{i_l} X_{t_{l-1},h}(s_1^{l-1},...,s_{m_{l-1}}^{l-1})|X_{t_{l-1}})\Bigg) \\
&= \mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot ... \cdot X_{t_{l-2},h}(s_1^{l-2},...,s_{m_{l-2}}^{l-2})\sum_{j_l=0}^{m_l}\sum_{i_l=0}^{j_l}\sum_{j_{l-1}=0}^{m_{l-1}+i_l}\sum_{i_{l-1}=0}^{j_{l-1}} L_h^2 X_{t_{l-1}}^{i_{l-1}}\right) \\
&= \mathbb{E}\left(\sum_{j_l=0}^{m_l}\sum_{i_l=0}^{j_l}\sum_{j_{l-1}=0}^{m_{l-1}+i_l}\sum_{i_{l-1}=0}^{j_{l-1}} ... \sum_{j_1=0}^{m_1+i_2}\sum_{i_1=0}^{j_1} L_h^l X_{t_1}^{i_1}\right) \\
&= \sum_{j_l=0}^{m_l}\sum_{i_l=0}^{j_l}\sum_{j_{l-1}=0}^{m_{l-1}+i_l}\sum_{i_{l-1}=0}^{j_{l-1}} ... \sum_{j_1=0}^{m_1+i_2}\sum_{i_1=0}^{j_1}\sum_{i_0=0}^{i_1} L_h^l X_0^{i_0}.
\end{aligned}$$

Again by Lemma 4.3.1, there exists a constant L, solely relies on m_1,...,m_l, such that $|L_h^l| \le Lh^l$. □

Theorem 4.3.4. *For any positive integer n, under assumptions 4.3.1 and 4.3.2, we* *have*

$$\mathbb{E}\left[\left(\sum_{i=1}^{T/h}(\eta X_{(i-1)h}+(1-\eta)X_{ih})h-\int_0^T V_s\mathrm{d}s\right)^{2n}\right]=O(h^{2n}).$$

Proof. Consider $t_i \in \{0,h,2h,...,T\}$, $i=1,2,...,l$, and we assume $t_1<t_2<...<t_l$, without the loss of generalisation. We let $\mathcal{N}$ be an element of the set $A_n \triangleq \{(m_1,m_2,...,m_l)|\sum_{i=1}^{l} m_i = 2n\}$, in which m_i, $i=1,...,l$ is a positive integer, and $1\le l\le 2n$. For instance, if $2n=2$, then $\mathcal{N}=(2)$ or $\mathcal{N}=(1,1)$. If $2n=4$, then the combinations of $\mathcal{N}$ are (4), $(3,1)$, $(1,3)$, $(2,2)$, $(2,1,1)$, $(1,2,1)$, $(1,1,2)$ and $(1,1,1,1)$. It is important to know that for any n, the number of combinations is finite. For any $\mathcal{N}=(m_1,m_2,...,m_l)$, we write $(t_1,...,t_l)\propto\mathcal{N}$ if there are m_1 number of t_1, m_2 number of t_2, and so forth.

By the binomial expansion and the mean value theorem, we have

$$\begin{aligned}
&\mathbb{E}\left(\sum_{i=1}^{T/h}(\eta X_{(i-1)h}+(1-\eta)X_{ih})h-\int_0^T X_s\mathrm{d}s\right)^{2n}\\
&=\sum_{\mathcal{N}\in A_n}\sum_{(t_1,...,t_l)\propto\mathcal{N}} C_{\mathcal{N}}\int_{t_1}^{t_1+h}...\int_{t_1}^{t_1+h}\int_{t_2}^{t_2+h}...\int_{t_2}^{t_2+h}...\int_{t_l}^{t_l+h}...\int_{t_l}^{t_l+h}\\
&\quad\mathbb{E}\left(X_{t_1,h}(s_1^1,...,s_{m_1}^1)\cdot...\cdot X_{t_l,h}(s_1^l,...,s_{m_l}^l)\right)\mathrm{d}s_1^1...\mathrm{d}s_{m_1}^1\mathrm{d}s_1^2...\mathrm{d}s_{m_2}^2..\mathrm{d}s_1^l...\mathrm{d}s_{m_l}^l\\
&=\sum_{\mathcal{N}\in A_n}\sum_{(t_1,...,t_l)\propto\mathcal{N}} C_{\mathcal{N}}h^{2n}\mathbb{E}\left(X_{t_1,h}(s_1^{1,*},...,s_{m_1}^{1,*})\cdot...\cdot X_{t_l,h}(s_1^{l,*},...,s_{m_l}^{l,*})\right)\\
&=\sum_{\mathcal{N}\in A_n}\sum_{(t_1,...,t_l)\propto\mathcal{N}} C_{\mathcal{N}}h^{2n}L_h^l,
\end{aligned}$$

where $C_{\mathcal{N}}$ is a constant with fixed $\mathcal{N}$, and $s_i^{j,*}$, for any $j=1,...,l$ and $i=1,...,m_j$, is a number in $[t_i,t_i+h]$, determined by the mean value theorem. The last equation is due to Lemma 4.3.2, and $|L_h^l|\le Lh^l$ with L a constant only relies on m_j, $j=1,...,l$, ie, only relies on $\mathcal{N}$.

Then, it suffices to prove

$$\sum_{(t_1,...,t_l)\propto\mathcal{N}} h^{2n}L_h^l = O(h^{2n})$$

for any combination $\mathcal{N}\in A_n$. We shall see that the number of combinations of such $(t_1,...,t_l)$ is $O((T/h)^l)$, as there are only l number of different t_i, $i=1,...,l$, and each t_i can take any of $O(T/h)$ values. Thus, the statement above is true. Since the number of the element $\mathcal{N} \in A_n$ is finite, the proof is complete. $\square$

Having proved Theorem 4.3.4 as expected, let us apply it to the Heston model by letting $X_t = V_t$, and we shall check whether the two assumptions are satisfied. The assumption 4.3.1 is valid due to (4.12). The other assumption 4.3.2 with respect to the trapezoidal rule is also true due to the lemma below.

Lemma 4.3.3. *Suppose V_t is given. Then, for the trapezoidal rule with $\eta = 1/2$, we have*

$$\lim_{h\to 0} \mathbb{E}\left[|V_{t,h}(s)|^m \,|V_t\right] = 0$$

for any non-negative integer m.

Proof. From the SDE of the variance process, we obtain

$$V_s - V_t = \int_t^s k(\theta - V_l)\mathrm{d}l + \int_t^s \sigma\sqrt{V_l}\mathrm{d}W_l,$$

where $t < s$. Thus, with substitution, we have

$$\mathbb{E}\left[|V_s - V_t|^m \,|V_\tau\right] \le 2^m\mathbb{E}\left[\left|\int_t^s k(\theta - V_l)\mathrm{d}l\right|^m \middle| V_\tau\right] + 2^m\mathbb{E}\left[\left|\int_t^s \sigma\sqrt{V_l}\mathrm{d}W_l\right|^m \middle| V_\tau\right],$$

where $\tau \in [0,t]$. It follows by the Cauchy-Schwartz inequality and the Burkholder-Davis-Gundy inequality that

$$\begin{aligned}\mathbb{E}\left[\left|\int_t^s \sqrt{V_l}\mathrm{d}W_l\right|^m \middle| V_\tau\right] &\le \mathbb{E}^{1/2}\left[\left(\int_t^s \sqrt{V_l}\mathrm{d}W_l\right)^{2m} \middle| V_\tau\right] \\ &\le c_m\mathbb{E}^{1/2}\left[\left(\int_t^s V_l\mathrm{d}l\right)^m \middle| V_\tau\right],\end{aligned}$$

where c_m here only depends on m. Furthermore, we have

$$\mathbb{E}\left[\left|\int_t^s k(\theta - V_l)\mathrm{d}l\right|^m \middle| V_\tau\right] \le (2k\theta)^m(s-t)^m + (2k)^m\mathbb{E}\left[\left(\int_t^s V_l\mathrm{d}l\right)^m \middle| V_\tau\right].$$

By the mean value theorem, we can see that

$$\mathbb{E}\left[\left(\int_t^s V_l \mathrm{d}l\right)^m \middle| V_\tau\right] = (s-t)^m \mathbb{E}\left(V_{l_1}...V_{l_m}|V_\tau\right)$$

with $l_1...l_m \in [t,s]$. It indicates that $\mathbb{E}\left[\left(\int_t^s V_l \mathrm{d}l\right)^m \middle| V_\tau\right]$ converges to zero when s goes to t. Therefore, we have

$$\lim_{h\to 0} \mathbb{E}\left[|V_s - V_t|^m |V_t\right] = 0,$$

where $s \in [t, t+h]$. The proof of

$$\lim_{h\to 0} \mathbb{E}\left[|V_s - V_{t+h}|^m |V_t\right] = 0$$

is analogous. We employ the inequality

$$\lim_{h\to 0} \mathbb{E}\left[|V_{t,h}(s)|^m |V_t\right] \leq \lim_{h\to 0} \mathbb{E}\left[|V_s - V_t|^m |V_t\right] + \mathbb{E}\left[|V_s - V_{t+h}|^m |V_t\right],$$

and the proof is complete. □

Then, the convergence theorem for the Heston model comes immediately.

Theorem 4.3.5. *For any positive integer n, we have*

$$\mathbb{E}\left[\left(\fint_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^{2n}\right] = O(h^{2n}).$$

Proof. By applying Theorem 4.3.4, with the assumptions satisfied due to (4.12) and Lemma 4.3.3, this theorem follows directly. □

Remark 4.3.3. *Theorem 4.3.4 applies to any quadrature rules, which satisfy assumption 4.3.2. One can check following Lemma 4.3.3 that assumption 4.3.2 still holds for Euler discretization. Therefore, we have*

$$\mathbb{E}\left[\left(\sum_{i=1}^{T/h} V_{(i-1)h} h - \int_0^T V_s \mathrm{d}s\right)^{2n}\right] = O\left(h^{2n}\right).$$

Lemma 4.3.4. (Dufresne). *For any $\tau \in \mathbb{R}$, and $T \in \mathbb{R}^+$, $\mathbb{E}\left[\left(\int_0^T V_s \mathrm{d}s\right)^\tau\right]$ is finite.*

Proof. See Theorem 4.1(a) in Dufresne (2001). □

Now, we can derive the convergence rate of the variance of the MLMC estimator, based on the results we have.

Theorem 4.3.6. *The convergence rate of the variance in the path-independent simulation is two, within the full parameter regime, and we have*

$$\mathrm{Var}(\ln \tilde{S}_{T,h} - \ln \tilde{S}_{T,Mh}) = O(h^2).$$

where M can be any positive integer.

Proof. We obtain

$$\begin{aligned}
\mathbb{E}\left[\left(\sqrt{\int\!\!\!\!-_0^T V_s^h \mathrm{d}s} - \sqrt{\int_0^T V_s \mathrm{d}s}\right)^2\right] &= \mathbb{E}\left[\left(\frac{\int\!\!\!\!-_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s}{\sqrt{\int\!\!\!\!-_0^T V_s^h \mathrm{d}s} + \sqrt{\int_0^T V_s \mathrm{d}s}}\right)^2\right] \\
&\le \mathbb{E}\left[\left(\frac{\int\!\!\!\!-_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s}{\sqrt{\int_0^T V_s \mathrm{d}s}}\right)^2\right] \\
&\le \mathbb{E}^{1/2}\left[\left(\int\!\!\!\!-_0^T V_s^h \mathrm{d}s - \int_0^T V_s \mathrm{d}s\right)^4\right] \mathbb{E}^{1/2}\left[\left(\frac{1}{\int_0^T V_s \mathrm{d}s}\right)^2\right],
\end{aligned}$$

where the last line is justified by the Cauchy-Schwarz inequality. Lemma 4.3.4 indicates

$$\mathbb{E}\left[\left(\frac{1}{\int_0^T V_s \mathrm{d}s}\right)^2\right] < \infty.$$

Therefore, with the aid of Theorem 4.3.5 and Theorem 4.3.2, we have

$$\mathbb{E}\left[\left(\ln \tilde{S}_{T,h} - \ln \tilde{S}_{T,Mh}\right)^2\right] = O(h^2),$$

Employing the equality

$$\mathrm{Var}(\ln \tilde{S}_{T,h} - \ln \tilde{S}_{T,Mh}) = \mathbb{E}\left[\left(\ln \tilde{S}_{T,h} - \ln \tilde{S}_{T,Mh}\right)^2\right] - \mathbb{E}^2\left(\ln \tilde{S}_{T,h} - \ln \tilde{S}_{T,Mh}\right),$$

and the inequality $\mathbb{E}^2(X) \leq \mathbb{E}\left[(X)^2\right]$ for any random variable X, we complete the proof. $\square$

Remark 4.3.4. *If we consider the Euler discretization instead, based on the following approximation*

$$\int_t^{t+h} V_s \mathrm{d}s \approx V_t h,$$

the convergence rate of the variance would be exactly the same. This can be proved in a analogous way as the we did in the theorem above for the trapezoidal rule. Although there is no improvement on the convergence rate of the variance when we use the trapezoidal rule for the Heston model, the main advantage is that the trapezoidal rule generally has a higher weak order than the Euler scheme. The MLMC prefers a numerical scheme with a higher weak order, as discussed.

4.3.3 Simulation for Path-dependent Options

The MLMC estimator for the path-dependent simulation is more complicated than that for the path-independent simulation, and we provide two MLMC estimators.

We denote by $\hat{S}_t^h$, $t = 0, h, 2h, ...T$, the approximation of the asset price S_t at the fine level with the step size h, and $\hat{S}_t^{Mh}$ be that at the coarse level with the step size Mh. At the fine level with the time step h, we let

$$\begin{aligned}\ln \hat{S}_{t+ih}^h = \ln \hat{S}_{t+(i-1)h}^h &+ \left(r - \frac{\rho k \theta}{\sigma}\right) h + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right) \frac{V_{t+(i-1)h} + V_{t+ih}}{2} h \\ &+ \frac{\rho}{\sigma}(V_{t+ih} - V_{t+(i-1)h}) + \sqrt{1-\rho^2}\sqrt{\frac{V_{t+(i-1)h} + V_{t+ih}}{2} h} N_i,\end{aligned}$$

for any $i = 1, ..., M$, and N_i are independent standard Normal random variables. On the other hand, at the coarse level with the time step Mh, we take

$$\begin{aligned}\ln \hat{S}_{t+Mh}^{Mh} = \ln \hat{S}_t^{Mh} &+ \left(r - \frac{\rho k \theta}{\sigma}\right) Mh + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right) \frac{V_t + V_{t+Mh}}{2} Mh \\ &+ \frac{\rho}{\sigma}(V_{t+Mh} - V_t) + \sqrt{1-\rho^2}\sqrt{\frac{V_t + V_{t+Mh}}{2} Mh} N.\end{aligned}$$

The key step in the MLMC is to construct the standard Normal variable N, based on the values of N_1, N_2,..., N_M simulated, in such a way that there is no additional bias introduced. That is, $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ and $\mathbb{E}(\hat{P}_{l+1} - \hat{P}_l)$ has the same expectation. There are a number of choices, and a simple approach might be

$$N = \frac{1}{\sqrt{M}}(N_1 + N_2 + ... + N_M). \tag{4.21}$$

This MLMC estimator is similar to the one used for the standard Euler and Milstein schemes. However, in the Heston model, where the variance process is simulated exactly, we have more flexibility to define an estimator. There is another estimator

$$N = \frac{\sum_{i=1}^{M} \sqrt{\int_{t+(i-1)h}^{t+ih} V_s^h \mathrm{d}s} N_i}{\sqrt{\int_t^{t+Mh} V_s^h \mathrm{d}s}}. \tag{4.22}$$

The numerical test in section 4.4 reveals that this estimator leads to a smaller variance $\mathrm{Var}(\ln \hat{S}_T^{Mh} - \ln \hat{S}_T^h)$ compared with the previous estimator. So, it is more suitable for the MLMC. We call the former (4.21) the standard estimator and the latter (4.22) the weighted average estimator.

For any path of $(V_s)_{s\in[0,T]}$ given, N calculated by (4.22) is a standard Normal random variable, due to that N_1,..., N_M simulated at the fine level are independent of the process $(V_s)_{s\in[0,T]}$. Since for all the path of the variance process, N has the same distribution, we shall say N is independent of the variance process. Then, $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ and in $\mathbb{E}(\hat{P}_{l+1} - \hat{P}_l)$ have the same expectation, which means the weighted average estimator is also free of additional bias. Analogous to the MLMC estimator for the path-independent simulation, both estimators here apply without any restriction on the type of the payoff function, as long as it is path-dependent.

Note that the variance process can reach zero, and in such a case, the weighted average estimator may be not well-defined due to the zero denominator. However, this is generally not a problem, because the probability that it happens is zero. Nevertheless, we shall be careful that the problem may arise if we consider a numerical scheme in which the approximated variance process, denoted by $(\widehat{V}_t)_{t=h,2h...,T}$, can be zero with a positive probability, such as the QE scheme. In this case, we can let the whole term $\sqrt{\frac{\widehat{V}_t+\widehat{V}_{t+Mh}}{2}}MhN$ be zero at the coarse level, without specifying the value of N. This treatment is applicable for the reason as follows. The zero denominator of the weighted average estimator implies zero $\widehat{V}_t$,$\widehat{V}_{t+h}$,....,$\widehat{V}_{t+Mh}$. This suggests that the diffusion terms $\sqrt{\frac{\widehat{V}_{t+(i-1)h}+\widehat{V}_{t+ih}}{2}hN_i}$, for $i = 0, ..., M$ at the fine level are all zeros. If we let the diffusion term $\sqrt{\frac{\widehat{V}_t+\widehat{V}_{t+Mh}}{2}MhN}$ at the coarse level also be zero, then there is no difference between the simulation of the diffusion terms at the fine level and at the coarse level, and the only difference comes from their drift terms. Since the MLMC estimator is constructed in such a way that we prefer a 'small' difference between the fine level and at the coarse level, this treatment solves the problem.

4.3.4 Convergence Analysis for Path-dependent Simulation

In our analysis, we shall focus on the weighted average estimator, because it is more efficient. Analogous to our analysis for the path-independent simulation, we analyse the convergence rate with respect to the logarithmic Heston price, which is

$$\mathbb{E}\max_{H=0,Mh,...,T}\left[\left(\ln\hat{S}_H^{Mh}-\ln\hat{S}_H^h\right)^2\right].$$

Lemma 4.3.5. *For any non-negative a_j, b_j, $j = 1, ..., n$, we have*

$$\left(\sqrt{\sum_{j=1}^n a_j}-\sqrt{\sum_{j=1}^n b_j}\right)^2 \le \sum_{j=1}^n\left(\sqrt{a_j}-\sqrt{b_j}\right)^2.$$

Proof. Due to the inequality $a_ib_j + a_jb_i \ge 2\sqrt{a_ia_jb_ib_j}$ for any non-negative a and b, we can get

$$\left(\sqrt{\sum_{j=1}^n a_j\sum_{j=1}^n b_j}\right)^2 \ge \left(\sum_{j=1}^n\sqrt{a_jb_j}\right)^2.$$

The lemma follows immediately. □

Theorem 4.3.7. *There is an inequality*

$$\mathbb{E}\max_{H=0,Mh,...,T}\left[\left(\ln\hat{S}_H^{Mh}-\ln\hat{S}_H^h\right)^2\right] \le E_d(h)+E_d(Mh),$$

where we define

$$
\begin{aligned}
E_d(\delta) = c\mathbb{E}\max_{H=0,Mh,\ldots,T}&\left[\left(⨍_0^H V_s^\delta \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right)^2\right] \\
&+ c\sum_{j=1}^{T/\delta}\mathbb{E}\left[\left(\sqrt{⨍_{(j-1)\delta}^{j\delta} V_s^\delta \mathrm{d}s} - \sqrt{\int_{(j-1)\delta}^{j\delta} V_s \mathrm{d}s}\right)^2\right].
\end{aligned}
$$

Proof. We have

$$
\begin{aligned}
&\mathbb{E}\max_{H=0,Mh,\ldots,T}\left[\left(\ln \hat{S}_H^{Mh} - \ln \hat{S}_H^h\right)^2\right] \\
&\leq 2\mathbb{E}\max_{H=0,Mh,\ldots,T}\left[\left(\ln \hat{S}_H^{Mh} - \ln S_H^{Mh}\right)^2\right] + 2\mathbb{E}\max_{H=0,Mh,\ldots,T}\left[\left(\ln \hat{S}_H^h - \ln S_H^{Mh}\right)^2\right]
\end{aligned}
\tag{4.23}
$$

where the exact solution $\ln S_H^{Mh} = \sum_{j=1}^{H/(Mh)}\left(\ln S_{jMh}^{Mh} - \ln S_{(j-1)Mh}^{Mh}\right) + \ln S_0^{Mh}$, $\ln S_0^{Mh} = \ln S_0$, and we have the recursion formula

$$
\begin{aligned}
\ln S_{t+Mh}^{Mh} = \ln S_t^{Mh} &+ \left(r - \frac{\rho k\theta}{\sigma}\right) Mh + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\int_t^{t+Mh} V_s \mathrm{d}s \\
&+ \frac{\rho}{\sigma}(V_{t+Mh} - V_t) + \sqrt{1-\rho^2}\sqrt{\int_t^{t+Mh} V_s \mathrm{d}s}N,
\end{aligned}
$$

where $t = 0, Mh, 2Mh, \ldots$, and N at each time-step is set to be the same random variable as N in the equation of $\ln \hat{S}_{t+Mh}^{Mh}$. We shall be careful that this definition of $\ln S_H^{Mh}$ relies on the step size Mh. It follows that

$$
\begin{aligned}
\ln \hat{S}_{t+Mh}^{Mh} - \ln S_{t+Mh}^{Mh} = \ln \hat{S}_t^{Mh} - \ln S_t^{Mh} &+ \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\left(⨍_t^{t+Mh} V_s^{Mh} \mathrm{d}s - \int_t^{t+Mh} V_s \mathrm{d}s\right) \\
&+ \sqrt{1-\rho^2}\left(\sqrt{\frac{V_t + V_{t+Mh}}{2}Mh} - \sqrt{\int_t^{t+Mh} V_s \mathrm{d}s}\right)N,
\end{aligned}
$$

and

$$
\begin{aligned}
&\ln \hat{S}^h_{t+Mh} - \ln S^{Mh}_{t+Mh} \\
&= \ln \hat{S}^h_t - \ln S^{Mh}_t + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\left[\sum_{i=1}^{M}\left(⨍_{t+(i-1)h}^{t+ih} V^h_s \mathrm{d}s\right) - \int_t^{t+Mh} V_s \mathrm{d}s\right] \\
&\quad + \sqrt{1-\rho^2}\left[\sum_{i=1}^{M}\left(\sqrt{\frac{V_{t+(i-1)h} + V_{t+ih}}{2}} hN_i\right) - \sqrt{\int_t^{t+Mh} V_s \mathrm{d}s N}\right] \\
&= \ln \hat{S}^h_t - \ln S^{Mh}_t + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\left(⨍_t^{t+Mh} V^h_s \mathrm{d}s - \int_t^{t+Mh} V_s \mathrm{d}s\right) \\
&\quad + \sqrt{1-\rho^2}\left(\sqrt{⨍_t^{t+Mh} V^h_s \mathrm{d}s} - \sqrt{\int_t^{t+Mh} V_s \mathrm{d}s}\right) N,
\end{aligned}
$$

where the last equality is due to the application of the weighted average estimator (4.22). The formulas above suggests that for each $t = jMh$, $j = 1, 2, ..., T/(Mh)$, there is an independent random variable N, which we rewrite as Z_j. We let

$$
Z = \frac{\sum_{j=1}^{H/(Mh)}\left(\sqrt{⨍_{(j-1)Mh}^{jMh} V^h_s \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right) Z_j}{\sqrt{\sum_{j=1}^{H/(Mh)}\left(\sqrt{⨍_{(j-1)Mh}^{jMh} V^h_s \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right)^2}},
$$

which is constructed in a similar way as the weighted average estimator, and thus Z is a standard Normal random variable, independent of the variance process. Therefore, we have

$$
\begin{aligned}
&\mathbb{E}\max_{H=0,Mh,\dots,T}\left[\left(\ln \hat{S}_H^h - \ln S_H^{Mh}\right)^2\right] \\
&= \mathbb{E}\max_H\left[\left(\sum_{j=1}^{H/(Mh)}\left((\ln \hat{S}_{jMh}^h - \ln \hat{S}_{(j-1)Mh}^h) - (\ln S_{jMh}^{Mh} - \ln S_{(j-1)Mh}^{Mh})\right)\right)^2\right] \\
&= \mathbb{E}\max_H\Bigg\{\Bigg[c\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right) \\
&\quad + c\sum_{j=1}^{H/(Mh)}\left(\sqrt{\fint_{(j-1)Mh}^{jMh} V_s^h \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right) Z_j\Bigg]^2\Bigg\} \\
&= \mathbb{E}\max_H\Bigg\{\Bigg[c\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right) \\
&\quad + c\sqrt{\sum_{j=1}^{H/(Mh)}\left(\sqrt{\fint_{(j-1)Mh}^{jMh} V_s^h \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right)^2} Z\Bigg]^2\Bigg\} \\
&\leq c\mathbb{E}\max_H\left[\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right)^2\right] \\
&\quad + c\sum_{j=1}^{T/(Mh)}\mathbb{E}\left[\left(\sqrt{\fint_{(j-1)Mh}^{jMh} V_s^h \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right)^2\right]\mathbb{E}(Z^2) \\
&\leq c\mathbb{E}\max_H\left[\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right)^2\right] \\
&\quad + c\sum_{j=1}^{T/h}\mathbb{E}\left[\left(\sqrt{\fint_{(j-1)h}^{jh} V_s^h \mathrm{d}s} - \sqrt{\int_{(j-1)h}^{jh} V_s \mathrm{d}s}\right)^2\right]\mathbb{E}(Z^2),
\end{aligned}
\tag{4.24}
$$

where the last inequality is due to the Lemma 4.3.5.

The analysis of $\mathbb{E}\max_{H=0,Mh,\dots,T}\left[\left(\ln \hat{S}_H^{Mh} - \ln S_H^{Mh}\right)^2\right]$ is similar but simpler, and we obtain

$$
\begin{aligned}
&\mathbb{E}\max_{H=0,Mh,\dots,T}\left[\left(\ln \hat{S}_H^{Mh} - \ln S_H^{Mh}\right)^2\right] \\
&\leq c\mathbb{E}\max_H\left[\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s\right)^2\right] \\
&\quad + c\sum_{j=1}^{T/(Mh)}\mathbb{E}\left[\left(\sqrt{\fint_{(j-1)Mh}^{jMh} V_s^{Mh} \mathrm{d}s} - \sqrt{\int_{(j-1)Mh}^{jMh} V_s \mathrm{d}s}\right)^2\right]\mathbb{E}(Z^2),
\end{aligned}
\tag{4.25}
$$

which is the same as the expression (4.24), except that we replace h with Mh.

Inequalities (4.24) and (4.25), together with (4.23) finish the proof. $\square$

Then, we shall analyse the bounds $E_d(h)$ and $E_d(Mh)$ in Theorem 4.3.7 for the convergence rate. We require the following result, which is available at several pieces of literature, for example, Dereich, Neuenkirch & Szpruch (2012): for $p > -\frac{2k\theta}{\sigma^2}$, we have

$$\sup_{t\in[0,T]} \mathbb{E}(V_t^p) < \infty.$$

Lemma 4.3.6. *Suppose V_t is given, then we have*

$$\mathbb{E}\left[\left(\int_t^{t+h} V_s \mathrm{d}s - \frac{V_t + V_{t+h}}{2}h \,\middle|\, V_t\right)^2\right] = \sum_{j=0}^{2} L_h^3 V_t^j,$$

where L_h^3 is independent of V_t, and there exists a constant L, such that $|L_h^3| \leq Lh^3$ for all h.

Proof. By the mean value theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_t^{t+h} V_s \mathrm{d}s - \frac{V_t + V_{t+h}}{2}h \,\middle|\, V_t\right)^2\right] \\
&= \int_t^{t+h}\int_t^{t+h} \mathbb{E}\left[\left(V_{s_1} - \frac{V_t + V_{t+h}}{2}\right)\left(V_{s_2} - \frac{V_t + V_{t+h}}{2}\right) \middle|\, V_t\right] \mathrm{d}s_1 \mathrm{d}s_2 \\
&= h^2\mathbb{E}\left[\left(V_{s_1^*} - \frac{V_t + V_{t+h}}{2}\right)\left(V_{s_2^*} - \frac{V_t + V_{t+h}}{2}\right) \middle|\, V_t\right],
\end{aligned}$$

where $s_1^*, s_2^* \in [t, t+h]$. We can apply Lemma 4.3.1 where the two assumptions are satisfied due to (4.12) and Lemma 4.3.3, and we can see that there exists a constant L, such that

$$\mathbb{E}\left[\left(V_{s_1^*} - \frac{V_t + V_{t+h}}{2}\right)\left(V_{s_2^*} - \frac{V_t + V_{t+h}}{2}\right) \middle|\, V_t\right] = \sum_{j=0}^{2} L_h^1 V_t^j,$$

where $|L_h^1| \leq Lh$ for all h. Therefore, it follows that

$$\mathbb{E}\left[\left(\int_t^{t+h} V_s \mathrm{d}s - \frac{V_t + V_{t+h}}{2}h \,\middle|\, V_t\right)^2\right] = \sum_{j=0}^{2} L_h^3 V_t^j,$$

where $|L_h^3| \leq Lh^3$ for all h. $\square$

Lemma 4.3.7. *The convergence rate of the drift part for the path-dependent simulation is one, and we have*

$$\mathbb{E} \max_{H=0,h,\dots,T} \left[\left(\fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s \right)^2 \right] = O(h).$$

Proof. We define

$$e_H \triangleq \fint_0^H V_s^h \mathrm{d}s - \int_0^H V_s \mathrm{d}s,$$

which represents the error at time H. We see the error is generated through the recursion

$$e_0 = 0,$$

$$e_{(l+1)h} = e_{lh} + \gamma_{lh},$$

for $l = 0, 1, \dots, H/h - 1$, where for any $t \in [0, T]$, and we have

$$\gamma_t = \frac{V_t + V_{t+h}}{2} - \int_t^{t+h} V_s \mathrm{d}s.$$

Multiplying both sides with $e_{(l+1)h}$, we obtain

$$e_{(l+1)h}^2 \leq \frac{1}{2} e_{(l+1)h}^2 + \frac{1}{2} e_{lh}^2 + e_{(l+1)h} \gamma_{lh},$$

and by the recursion it follows that

$$e_H^2 \leq 2 \sum_{l=0}^{H/h-1} e_{(l+1)h} \gamma_{lh},$$

for any $H = h, 2h, \dots, T$. Then, we have

$$\max_{H=0,h,\dots,T} e_H^2 \leq 2 \sum_{l=0}^{T/h-1} |e_{(l+1)h} \gamma_{lh}|,$$

and it yields

$$\max_{H=0,h,\dots,T} |e_H| \leq 2 \sum_{l=0}^{T/h-1} |\gamma_{lh}|. \tag{4.26}$$

Furthermore, by Lemma 4.3.6, we obtain

$$\mathbb{E}(\gamma_t^2) = \sum_{j=0}^{2} L_h^3 \mathbb{E}(V_t^j),$$

where $|L_h^3| \leq Lh^3$ with L a constant. Since $\mathbb{E}(V_t^j)$ is bounded on $t \in [0,T]$, there is a constant L^*, such that

$$\mathbb{E}(\gamma_t^2) \leq L^* h^3.$$

Using (4.26) and Minkowski's inequality on $\mathbb{L}^2$ space, we can get

$$\mathbb{E} \max_{H=0,h,\ldots,T} e_H^2 \leq \mathbb{E}\left[\left(2\sum_{l=0}^{T/h-1} |\gamma_{lh}|\right)^2\right] \leq \left(2\sum_{l=0}^{T/h-1} \sqrt{\mathbb{E}(\gamma_{lh})^2}\right)^2 = O(h).$$

The proof is complete. □

Remark 4.3.5. *The link between the global error and the local error in the proof of Lemma 4.3.7 is from Proposition 3.3 of Dereich et al. (2012).*

Lemma 4.3.8. *For any non-negative X and Y, there is an inequality*

$$\left(\sqrt{X} - \sqrt{Y}\right)^2 \leq |X - Y|.$$

Proof. Since $X + Y \geq 2\sqrt{XY}$, we have

$$\left(\sqrt{X} - \sqrt{Y}\right)^4 - (X - Y)^2 = 8XY - 4(X + Y)\sqrt{XY} \leq 0.$$

The lemma follows immediately. □

Based on the results above, we are able to provide the theorem below with a theoretical convergence rate for the Heston model.

Theorem 4.3.8. *The convergence rate of variance for the path-dependent simulation is explicitly half, within the full parameter regime, and we have*

$$\mathrm{Var} \max_{H=0,Mh,\ldots,T} \left| \ln \hat{S}_H^h - \ln \hat{S}_H^{Mh} \right| = O(h^{1/2}),$$

where M can be any positive integer.

Proof. Employing Lemma 4.3.8 and the Cauchy-Schwarz inequality, we can get

$$\begin{aligned}\mathbb{E}\left[\left(\sqrt{\frac{V_t+V_{t+h}}{2}h}-\sqrt{\int_t^{t+h}V_s\mathrm{d}s}\right)^2\right] &\leq \mathbb{E}\left|\frac{V_t+V_{t+h}}{2}h-\int_t^{t+h}V_s\mathrm{d}s\right| \\ &\leq \mathbb{E}^{1/2}\left[\left(\frac{V_t+V_{t+h}}{2}h-\int_t^{t+h}V_s\mathrm{d}s\right)^2\right] \\ &= \left(\sum_{j=0}^{2}L_h^3\mathbb{E}(V_t^j)\right)^{1/2},\end{aligned}$$

where $|L_h^3|\leq Lh^3$, and L is a constant. Due to the uniform boundedness of $\mathbb{E}(V_t^j)$ on $t\in[0,T]$, we can see

$$\sup_{t\in[0,T]}\mathbb{E}\left[\left(\sqrt{\frac{V_t+V_{t+h}}{2}h}-\sqrt{\int_t^{t+h}V_s\mathrm{d}s}\right)^2\right]=O(h^{3/2}).$$

It follows that

$$\sum_{j=1}^{T/h}\mathbb{E}\left[\left(\sqrt{\int_{(j-1)h}^{jh}V_s^h\mathrm{d}s}-\sqrt{\int_{(j-1)h}^{jh}V_s\mathrm{d}s}\right)^2\right]=O(h^{1/2}).$$

Then, with the aid of Theorem 4.3.7 and Lemma 4.3.7, we prove that

$$\mathbb{E}\max_{H=0,Mh,\ldots,T}\left[\left(\ln\hat{S}_H^h-\ln\hat{S}_H^{Mh}\right)^2\right]=O(h^{1/2}).$$

Finally, we apply the inequality

$$\begin{aligned}&\mathrm{Var}\max_{H=0,Mh,\ldots,T}\left|\ln\hat{S}_H^h-\ln\hat{S}_H^{Mh}\right| \\ &=\mathrm{E}\max_{H=0,Mh,\ldots,T}\left[(\ln\hat{S}_H^h-\ln\hat{S}_H^{Mh})^2\right]-\mathbb{E}^2\max_{H=0,Mh,\ldots,T}\left|\ln\hat{S}_H^h-\ln\hat{S}_H^{Mh}\right|,\end{aligned}$$

and the inequality $\mathbb{E}^2(|X|)\leq\mathbb{E}\left[(X)^2\right]$ for any random variable X. The proof is complete. $\square$

Remark 4.3.6. *The convergence result on*

$$\mathrm{Var}\left(\max_H \left|\ln \hat{S}_H^h - \ln \hat{S}_H^{Mh}\right|\right)$$

is more informative than that on

$$\mathrm{Var}\left(\ln \hat{S}_T^h - \ln \hat{S}_T^{Mh}\right),$$

the latter of which appears in the majority of literature. This is because the former indicates the rate of

$$\mathbb{E}\left(\max_H \left[\left(\ln \hat{S}_H^h - \ln \hat{S}_H^{Mh}\right)^2\right]\right)$$

which is no smaller than

$$\mathbb{E}\left[\left(\ln \hat{S}_T^h - \ln \hat{S}_T^{Mh}\right)^2\right].$$

Remark 4.3.7. *Analogous to the path-independent simulation, if we consider the Euler discretization instead of the trapezoidal discretization, the convergence rate of variance of the MLMC estimator would be exactly the same. This result fills a gap in Altmayer & Neuenkirch (2015), when the zero boundary of the variance process is attainable.*

If the parameters of the variance process satisfies $\frac{2k\theta}{\sigma^2} > 1$, which is equivalent to the case where the zero boundary is never attainable, then, we have a higher theoretical convergence rate one. The Euler version of this result can be found at Altmayer & Neuenkirch (2015), but our proof is quite different.

Theorem 4.3.9. *Suppose $\frac{2k\theta}{\sigma^2} > 1$, then the convergence rate of variance for the path-dependent simulation is explicitly one, and we have*

$$\mathrm{Var} \max_{H=0,Mh,\ldots,T} \left|\ln \hat{S}_H^h - \ln \hat{S}_H^{Mh}\right| = O(h),$$

where M can be any positive integer.

Proof. For any $t \in [0, T]$, we see that

$$\mathbb{E}\left[\left(\sqrt{\frac{V_t + V_{t+h}}{2}h} - \sqrt{\int_t^{t+h} V_s \mathrm{d}s}\right)^2\right] = \mathbb{E}\left[\left(\frac{\frac{V_t+V_{t+h}}{2}h - \int_t^{t+h} V_s \mathrm{d}s}{\sqrt{\frac{V_t+V_{t+h}}{2}h} + \sqrt{\int_t^{t+h} V_s \mathrm{d}s}}\right)^2\right]$$
$$\leq \frac{2}{h}\mathbb{E}\left(\frac{\mathbb{E}\left[\left(\frac{V_t+V_{t+h}}{2}h - \int_t^{t+h} V_s \mathrm{d}s | V_t\right)^2\right]}{V_t}\right)$$
$$= \frac{2}{h}\sum_{j=0}^{2} L_h^3 \mathbb{E}(V_t^{j-1}),$$

where the last equality follows by Lemma 4.3.6, and we have $|L_h^3| \leq Lh^3$ with L a constant. We notice that there is a term $\mathbb{E}\left(\frac{1}{V_t}\right)$. As $\frac{2k\theta}{\sigma^2} > 1$ implies $\sup_{t\in[0,T]} \mathbb{E}\left(\frac{1}{V_t}\right) < \infty$, it follows

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left(\sqrt{\frac{V_t + V_{t+h}}{2}h} - \sqrt{\int_t^{t+h} V_s \mathrm{d}s}\right)^2\right] = O(h^2),$$

and therefore, we have

$$\sum_{j=1}^{T/h} \mathbb{E}\left[\left(\sqrt{\fint_{(j-1)h}^{jh} V_s^h \mathrm{d}s} - \sqrt{\int_{(j-1)h}^{jh} V_s \mathrm{d}s}\right)^2\right] = O(h).$$

With the aid of Theorem 4.3.7 and Lemma 4.3.7, we obtain

$$\mathbb{E} \max_{H=0,Mh,\ldots,T}\left[\left(\ln \hat{S}_H^h - \ln \hat{S}_H^{Mh}\right)^2\right] = O(h).$$

By the same analysis on the variance Var(.) as in Theorem 4.3.8, the proof is complete. □

4.3.5 QE MLMC

So far, we have established the MLMC estimators of both the path-independent and the path-dependent simulations for the Heston model, and the estimators are built based on the assumption that the variance process $(V_t)_{t\geq 0}$ can be exactly simulated.

Except for the exact methods to simulate the variance process exactly, for instance, Marsaglia & Tsang (2000), our MLMC technique for the Heston model is expected to work well in principle when implemented with the almost-exact methods such as Van Haastrecht & Pelsser (2010) and Malham & Wiese (2013). However, the QE scheme by Andersen (2008) is a biased scheme. Thus, there would be additional bias introduced when applying this MLMC technique, because $\hat{P}_l$ in $\mathbb{E}(\hat{P}_l - \hat{P}_{l-1})$ and $\mathbb{E}(\hat{P}_{l+1} - \hat{P}_l)$ might not have the same expectation. However, it is possible for the QE MLMC to yield a good numerical result, under certain circumstance. As explained in Andersen (2008), also in Van Haastrecht & Pelsser (2010), page 24-25, the bias is small for a small discretization size h, which may lead to a high value of the non-centrality parameter

$$\lambda = \frac{4ke^{-kh}}{\sigma^2(1-e^{-kh})}V_u = \frac{4k}{\sigma^2}\frac{1}{e^{kh}-1}V_u,$$

conditioned on V_u known. We prefer a large λ because the probability function of a non-central chi-squared random variable can be well represented by a power function applied to a Gaussian variable when λ is large. Therefore, in order to reduce the bias from the QE scheme, it would be more efficient to start MLMC at a level with multiple steps than to start it from a single step, if the maturity T is large.

4.4 Numerical Results

In this section, we will carry out three numerical experiments. In the first experiment, we numerically evaluate the convergence rates of the variance for the MLMC estimators, and compare them with the theoretical rates. In the second experiment, we compare our MLMC for the Heston model with the underlying scheme without MLMC, to test whether the computational saving is significant. In the last experiment, we extend the numerical analysis for the Asian option, which has a relatively complicated payoff function.

All the experiments here are conducted in Matlab 2014(a), and we use the function 'ncx2rnd' to exactly simulate the non-central chi-square random variable of the variance process. Despite this technique is not favoured in financial applications,

Table 4.1: Model Parameters

	Case I	Case II	Case III	Case IV
k	2	0.5	0.3	6.2
θ	0.09	0.04	0.04	0.02
σ	1	1	0.9	0.6
ρ	-0.3	-0.9	-0.5	-0.7
r	0.05	0	0	0

Note: In all cases, $T = 1$, $V(0) = \theta$, and $S(0) = 100$

it is a standard method for numerical tests due to its high accuracy. Our MLMC technique is suitable for all numerical methods that can simulate the variance process exactly or almost-exactly. Four sets of parameters are employed as in Table 4.1, all of which are interesting to the practice. Specifically, Case I is from Broadie & Kaya (2006) for equity options. Case II is for long-dated FX options, and Case III for long-term interest rate options as mentioned in Andersen (2008). The last Case IV comes from Glasserman & Kim (2011) representing for S&P 500 index options. Here, a unit of T is for one year, and for computational convenience, we let $T = 1$ for all. The strike of the option is set to be equal to the initial value of the asset S_0, and so, the option is called at-the-money. The exact prices can be calculated with a very high accuracy, using the technique by Kahl & Jäckel (2005), and they are 13.1365, 4.4034, 5.0997, 5.2774 respectively. Recall that when $2k\theta \geq \sigma^2$, the boundary zero of the variance process is unattainable, and when $2k\theta < \sigma^2$, the origin is attainable and reflecting, and the simulation concerning the latter is more challenging. As it is easy to verify, all the cases in the table fall into the latter category.

4.4.1 Numerical Convergence Rate

In the first experiment, we plot the variance

$$\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$$

to see how fast it converges to zero as the level l grows. Here, $\hat{P}_l$ is the price of the standard European call option approximated by the stochastic trapezoidal discretization with the step size equals to M^{-l}, corresponding to the level l. We let $M = 4$ for all the cases, regardless of whether they are path-independent or path-dependent. In our previous theoretical analysis, we have focused on options with bounded Lipschitz continuous payoff functions. The convergence rate of variance is 2 for the path-independent simulation in all parameter regimes. It is 1/2 for the path-dependent simulation in all parameter regimes, and it is 1 when $2k\theta \geq \sigma^2$. However, in this test, we consider options with unbounded Lipschitz continuous payoffs, which is exactly the case of the standard European call option. The number of samples we collect at each level is 1 million, so the standard deviation of the estimator for $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ is quite small.

For the path-independent simulation, the result is shown in Figure 4.1, which plots the logarithm of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ (the blue line) with the base M versus different levels l. For comparison, we also plot the logarithm of $\mathrm{Var}(\hat{P}_l)$ (the red line). As can be observed in all of the cases, $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ converges at rate 2, which is consistent with our theoretical analysis. Further, when $l = 4$, $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ is generally less than 4^{-6} of $\mathrm{Var}(\hat{P}_l)$.

For the path-dependent cases, we have two MLMC estimators: the standard estimator and the weighted average estimator, with the former plotted in black, and the latter plotted in blue, as displayed in Figure 4.2. They differ in the way to construct the Brownian motion at the coarse level using the Brownian motions simulated at the fine level. The same as the treatment of path independent cases, we illustrate the logarithm of $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ and $\mathrm{Var}(\hat{P}_l)$ with the base M at the different level l. The numerical result shows that in all the cases, the variances converge at rate 1, although in the theoretical analysis, there is a condition $2k\theta \geq \sigma^2$ imposed to guarantee the first order convergence. In addition, the weighted average estimator is much better than the standard estimator, because the implicit coefficient of the leading order term of the former is significantly smaller, despite that there is no improvement over the convergence rate.

To be rigorous, we repeat the experiments for all the cases only by changing $M = 2$, and the numerical convergence rates are the same as these when $M = 4$,

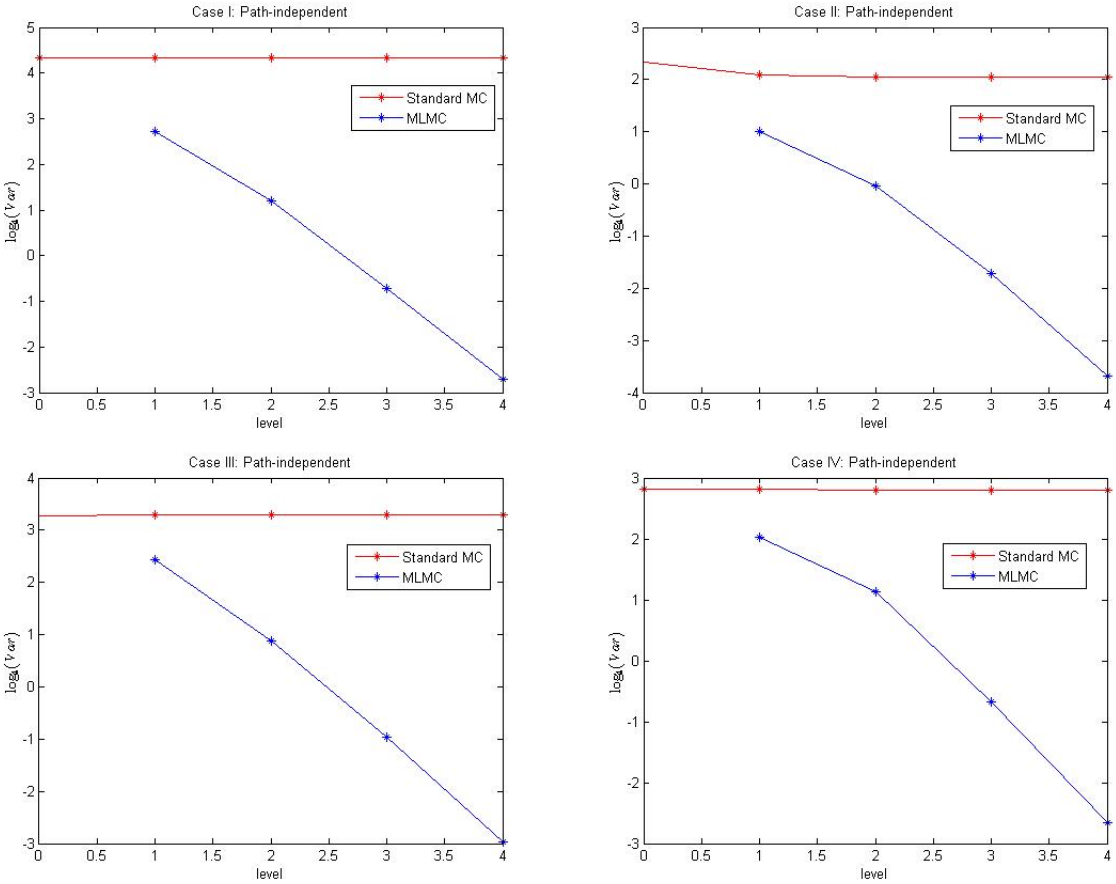

Figure 4.1: The comparison of $\mathrm{Var}(\hat{P}_l)$ and $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-independent cases I-IV, M = 4. The red line is for the standard MC, and the blue line is for the MLMC.

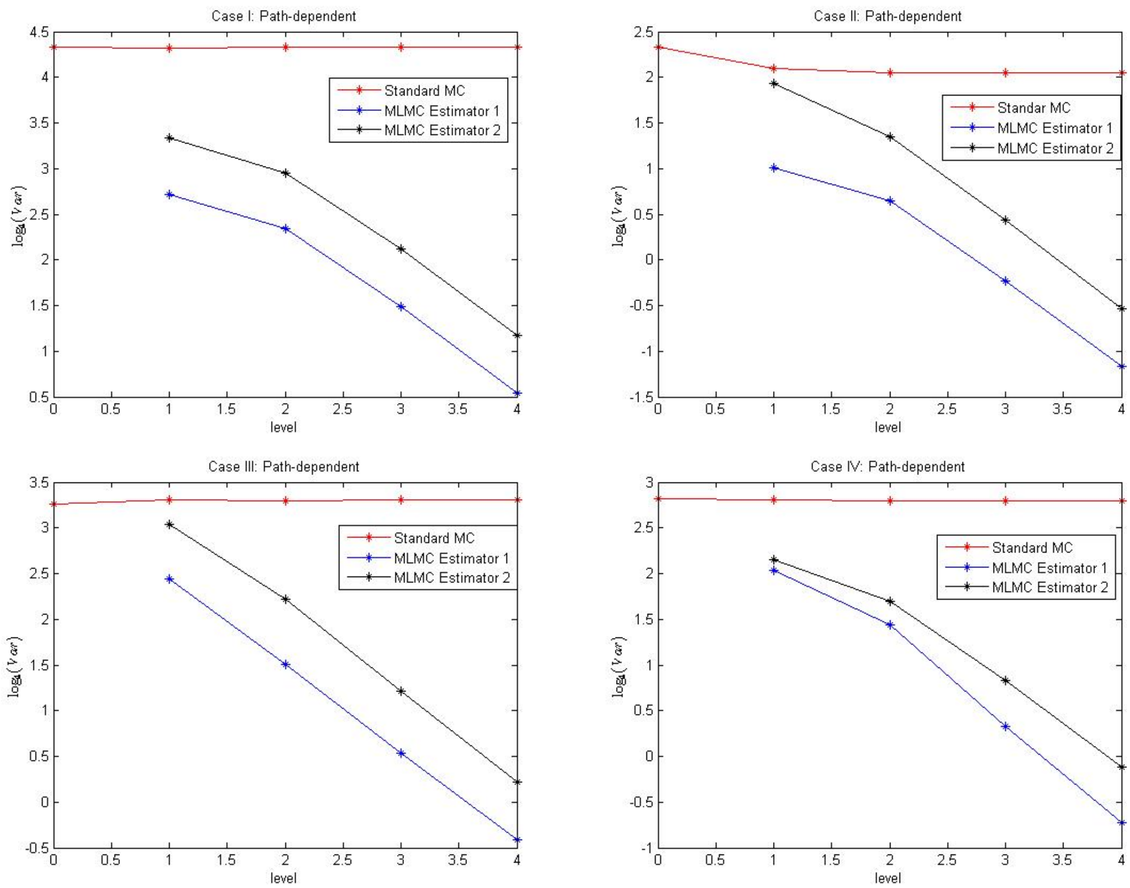

Figure 4.2: The comparison of $\mathrm{Var}(\hat{P}_l)$ and $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-dependent cases I-IV, M = 4. The blue plot is for the weighted average estimator and the black plot is for the standard estimator.

for the path-independent and the path-dependent simulations respectively. The results are demonstrated in Figure 4.3 and Figure 4.4. This means the convergence rate is independent of the choice of M. The result is consistent with our previous theoretical analysis, which does not rely on M either.

On the other hand, one shall notice that without the MLMC, the approximated asset prices at maturity T by the path-independent and path-dependent simulations have the same distribution function. This is because the path-dependent simulation is based on

$$\begin{aligned}\ln \hat{S}_T^h &= \ln \hat{S}_0 + \left(r - \frac{\rho k\theta}{\sigma}\right)T + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\sum_{i=1}^{T/h} \frac{V_{(i-1)h} + V_{ih}}{2}h \\ &\quad + \frac{\rho}{\sigma}(V_T - V_0) + \sum_{i=1}^{T/h} \sqrt{1-\rho^2}\sqrt{\frac{V_{(i-1)h} + V_{ih}}{2}h}N_i \\ &= \ln \hat{S}_0 + \left(r - \frac{\rho k\theta}{\sigma}\right)T + \left(\frac{\rho k}{\sigma} - \frac{1}{2}\right)\sum_{i=1}^{T/h} \frac{V_{(i-1)h} + V_{ih}}{2}h \\ &\quad + \frac{\rho}{\sigma}(V_T - V_0) + \sqrt{1-\rho^2}\sqrt{\sum_{i=1}^{T/h}\left(\frac{V_{(i-1)h} + V_{ih}}{2}h\right)}N,\end{aligned}$$

where N_i for any i is a standard Normal random variable, and

$$N = \sum_{i=1}^{T/h} \sqrt{\frac{V_{(i-1)h} + V_{ih}}{2}h}N_i \Bigg/ \sqrt{\sum_{i=1}^{T/h}\left(\frac{V_{(i-1)h} + V_{ih}}{2}h\right)}.$$

is again standard Normal distributed. The last equality of $\ln \hat{S}_T^h$ is just the discretization of $\ln \tilde{S}_{T,h}$ for the path-independent simulation. Therefore, there is no difference in the variance of the approximated option prices on the same level by the path-independent simulation and by the path-dependent simulation. This phenomenon can be observed by comparing Figure 4.1 and Figure 4.2, or by comparing Figure 4.3 and Figure 4.4.

In conclusion, the numerical results show that in all the cases we consider for the standard European call option, the convergence of variance is 2 for path-independent simulation and it is 1 for path-dependent simulation. Further, the choice of M does not affect the convergence rate.

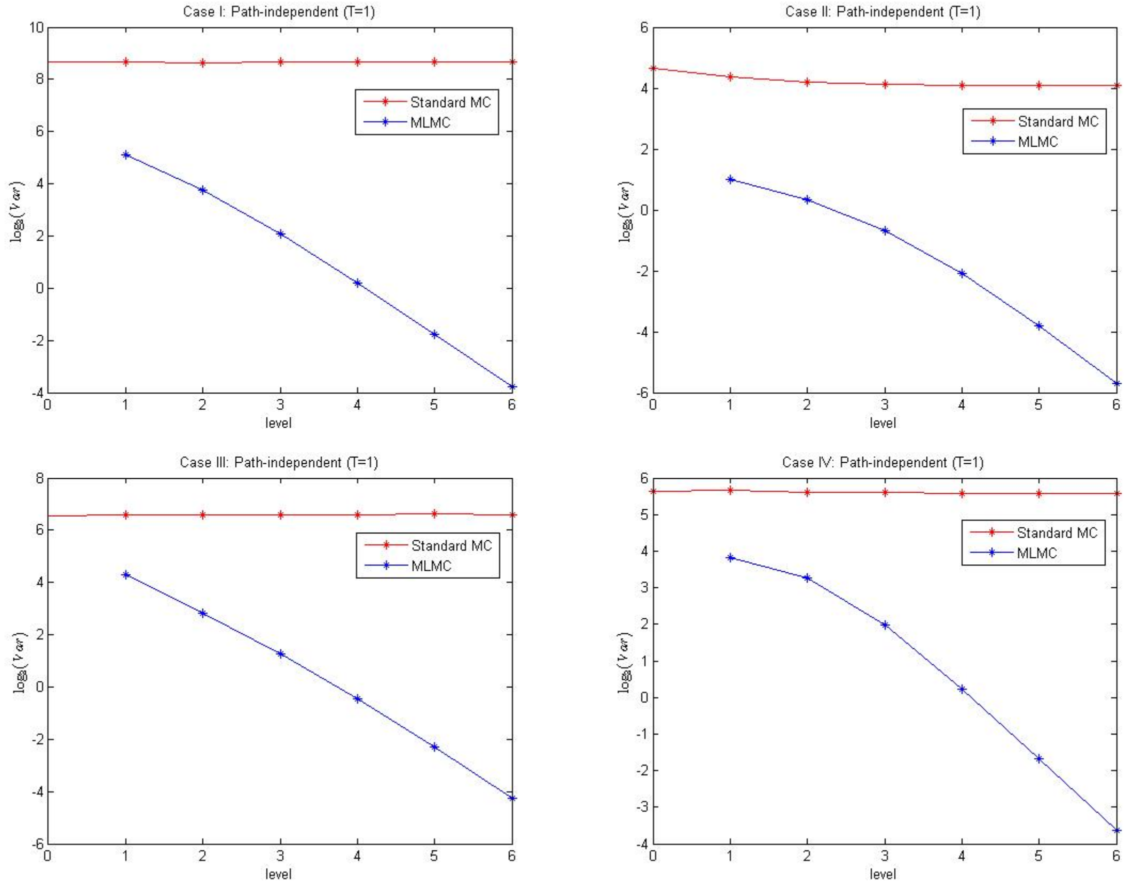

Figure 4.3: The comparison of $\mathrm{Var}(\hat{P}_l)$ and $\mathrm{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-independent cases I-IV, M $= 2$. The red line is for the standard MC, and the blue line is for the MLMC.

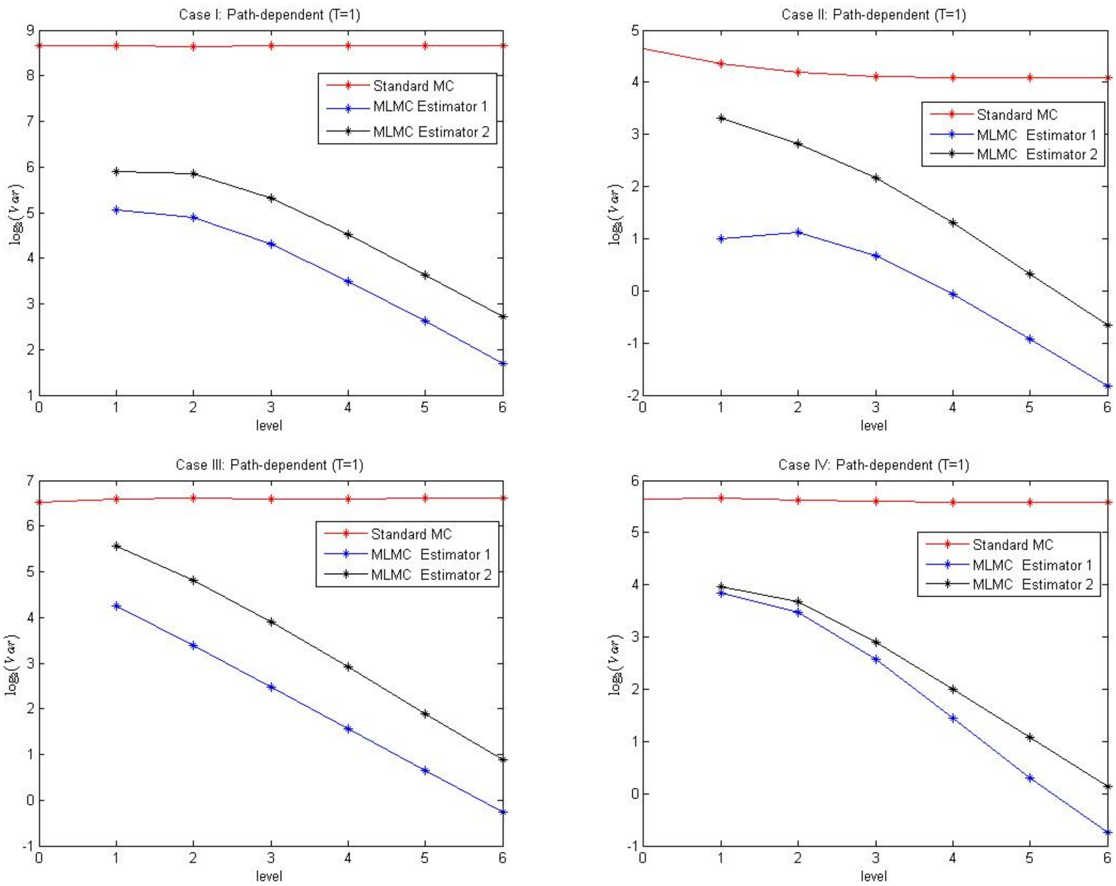

Figure 4.4: The comparison of $\text{Var}(\hat{P}_l)$ and $\text{Var}(\hat{P}_l - \hat{P}_{l-1})$ for path-dependent cases I-IV, M = 2. The blue plot is for the weighted average estimator and the black plot is for the standard estimator.

4.4.2 MLMC Performance

In this experiment, we shall pay our attention to the efficiency of our MLMC on the stochastic trapezoidal discretization for the Heston model compared with the underlying scheme without MLMC.

As was discussed previously, the accuracy of the standard Monte Carlo is determined by the step size and the number of the path simulations, while for MLMC, there is only one user-specified parameter ε, the square of which is approximately the mean squared error. Further, recall that in MLMC, we deal with the path-independent and path-dependent simulations separately using different MLMC estimators, so the accuracy and the computational complexity are expected to be quite different. We use the previous definition of the computational cost for the MLMC as discussed in Section 4.2, and the computational cost is

$$C = N_0 + \sum_{l=1}^{L} N_l (M^l + M^{l-1})$$

where N_l is the number of simulations specified by ε and l. For the standard Monte Carlo, we calculate it as

$$C^* = \sum_{l=0}^{L} N_l^* M^l$$

where $N_l^* = 2\xi^{-2} V(P_l)$, so the variance of the estimator is the same as it is in MLMC. This treatment is the same as what was used in Giles (2008*b*) for computational complexity.

In the numerical test, we let $M = 4$, and the weak convergence rate used in the MLMC algorithm to control the bias is set to be two. We compare the computational complexity of MLMC with that of the standard Monte Carlo. The result is illustrated in Figure 4.5, which plots $C\xi^2$ and $C^*\xi^2$ versus ξ. For the path-dependent simulation, we only consider the weighted average estimator because it is better than the standard estimator, as discussed in the previous experiment. From Figure 4.5, the computational saving by MLMC is significant in all the cases, both for path-independent and the path-dependent simulations, and the ratio is up to 10. As expected, the path-independent simulation has higher computational savings than that of the path-dependent simulation in general. For instance, when $\xi = 0.005$ in Case III, the computational saving is 7.9 for path-independent simulation, while it is 5.1 for path-dependent simulation. This is mainly because the MLMC estimator of path-independence has a higher convergence rate, and thus the variance is much smaller at a higher level, which leads to the smaller number of samples from the MLMC algorithm.

Next, we analyse the computational complexity in a more theoretical perspective. For a weak-order-two numerical scheme, the standard Monte Carlo computational cost would be $O(\xi^{-5/2})$, in order to produce a mean squared error $O(\xi^2)$. This is due to the asymptotic formula $\text{MSE}(\hat{e}) = O(n^{-1}) + O(h^{2\beta})$ we have discussed in Chapter two, where n is the number of samples, and h is the step size of the underlying numerical scheme. To obtain a MSE $O(\xi^2)$, we need to set $n = O(\xi^{-2})$ and $h = O(\xi^{1/2})$, and so the overall cost would be $nh^{-1} = O(\xi^{-5/2})$. As can be observed, the red plots show the Monte Carlo computational complexity is $O(\xi^{-5/2})$, which corresponds to the ratio 1/2 due to $O(\xi^{-5/2})\xi^2 = O(\xi^{-1/2})$. In terms of the the computational complexity for the MLMC, the theorem by Giles states that the computational cost is $O(\xi^{-2})$ for a MLMC estimator with a variance of second-order convergence, and it is $O(\xi^{-2}(\ln \xi)^2)$ for that with a variance of first-order convergence. Again, the plots on the left illustrate the numerical computational cost for the path-independent simulation is slightly higher than $O(\xi^{-2})$, and for the path-dependent simulation, it is roughly $O(\xi^{-2}(\ln \xi)^2)$. This means the computational cost in the plots below is in general consistent with what the theory predicts. The inconsistency can happen because there is no guarantee for the MLMC algorithm to provide a MSE of the order expected, as discussed in Section 4.2. Furthermore, in the first two pictures in a row, the plots are flat because the maximum level used in the MLMC is always two, even if we input a very small ξ. This is possibly due to the control (4.6) in the MLMC algorithm fails to work, for this particular set of parameters of the Heston model.

On the other hand, we notice that except for the weak and strong convergence rate, the performance of the MLMC also relies on the value of the parameter M

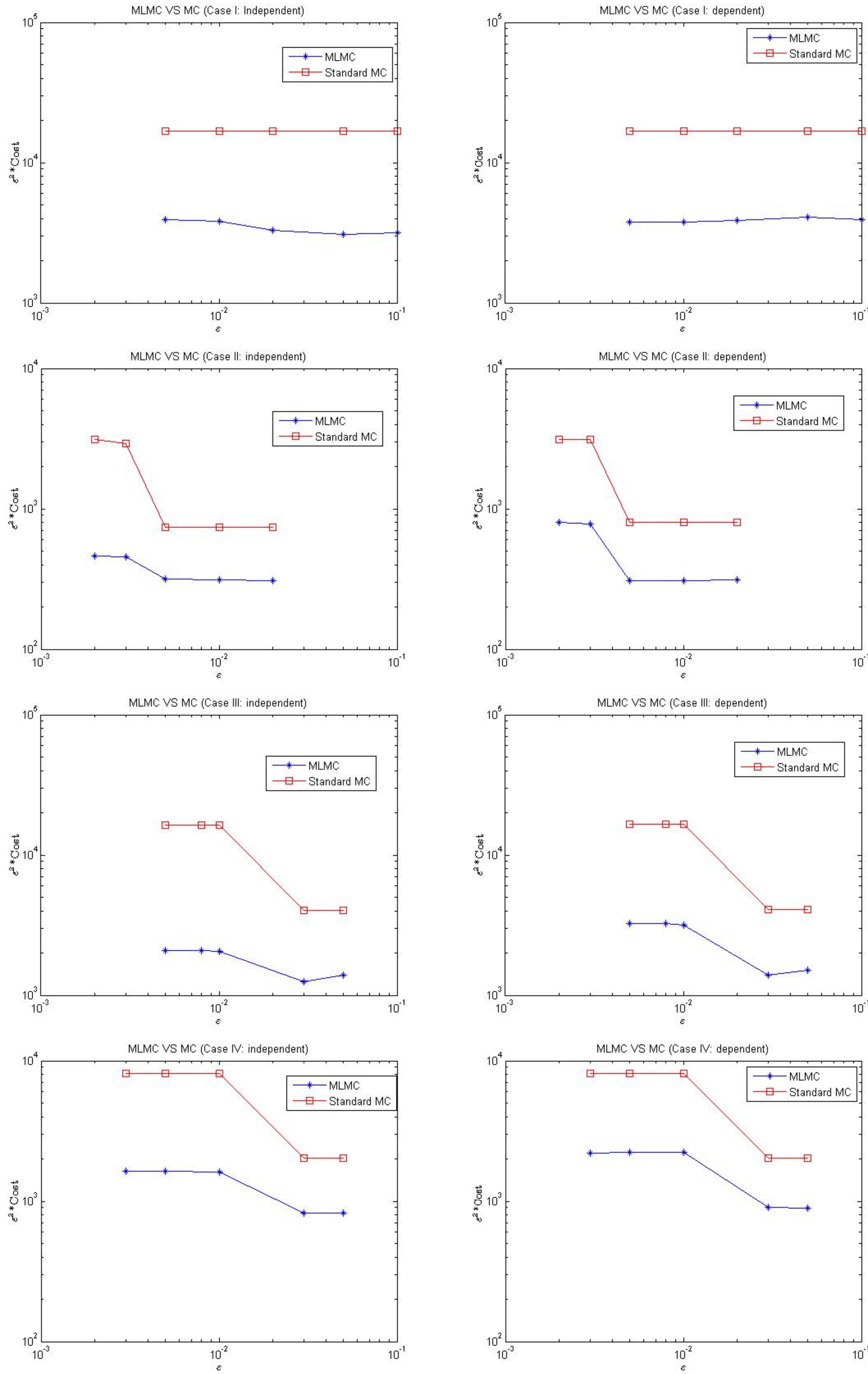

Figure 4.5: The comparison of the computational complexity between the MLMC and the standard MC. The left plots are for the path-independent simulation, and the right are for the path-dependent simulation.

Table 4.2: Parameters for Asian Option

k	1.0407	r	0
θ	0.0586	T	1
σ	0.5196	$V(0)$	0.0194
ρ	-0.6747	$S(0)$	100

in the MLMC algorithm. The value of M we use for the tests is not optimal. It is possible the performance of the MLMC would be significantly improved with a better choice of M.

4.4.3 Result for Asian Option

Finally, we consider the Asian option. Recall that the price of an Asian call option is

$$\mathbb{E}\left(\left(\frac{1}{T}\int_0^T S_t \mathrm{d}t - L\right)^+\right),$$

where $(S_t)_{t\in[0,T]}$ is the asset process, T is the maturity of the underlying option, and K is the strike. Here, we assume $(S_t)_{t\in[0,T]}$ is the Heston solution. When we use the Monte Carlo simulation for the numerical solution, it is convenient to approximate the integral

$$\int_0^T S_t \mathrm{d}t \approx \sum_{i=1}^{T/\Delta} \frac{S_{(i-1)\Delta} + S_{i\Delta}}{2}\Delta$$

based on the trapezoidal rule with step size Δ. For the MLMC, at the coarse level, we let $\Delta = Mh$, and at the fine level, we let $\Delta = h$. To price such an Asian option, we need the asset process before maturity T. Thus, only the path-dependent simulation can be used.

In this experiment, the parameters, available in Table 4.4.3, are from Smith (2007). Again, the zero boundary of the variance process is attainable, and we set $T = 1$ for computational convenience, together with $M = 4$. We investigate the convergence rate of variance, and analyse the MLMC complexity, as we did for the standard European call option. The result is illustrated in Figure 4.6, with the left side the plot of the variance and the right side the plot of the computational complexity. As can be observed from the left plot, the convergence rate of variance

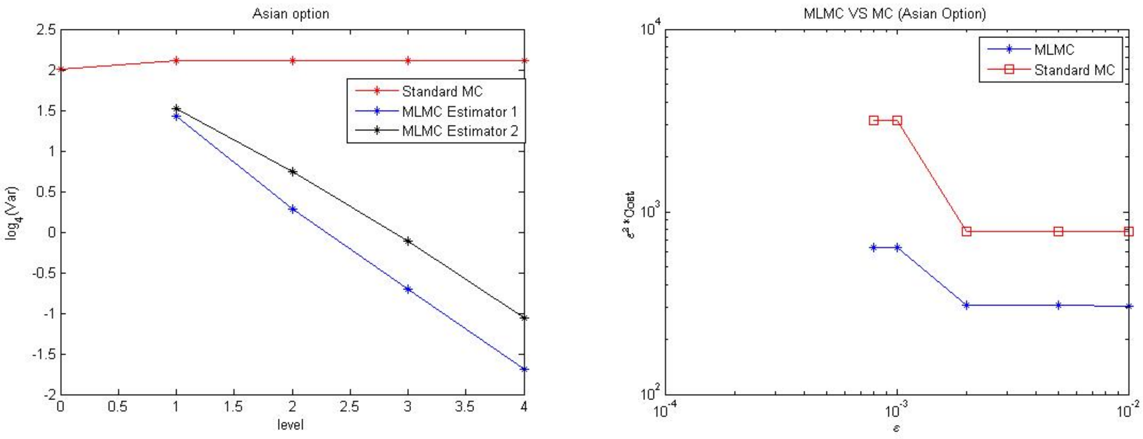

Figure 4.6: Numerical test on Asian option. The left is for the variance and the right is for the computational complexity. The blue plot is for the weighted average estimator and the black plot is for the standard estimator.

is one, and the weighted average estimator is significantly better than the standard estimator, which is consistent with our previous numerical analysis. On the other hand, the right plot shows that the computational saving from the MLMC is also significant, and it looks roughly $O(\xi^{-2}(\ln \xi)^2)$ against the standard Monte Carlo $O(\xi^{-5/2})$. It can also be witnessed that the MLMC can be 5 times more efficient.

Furthermore, since the Asian option is readily the expectation of the average asset price over $[0, T]$, the variance is generally smaller than that of the standard European option with similar values of parameters. In our case, the variance of Asian option priced by standard Monte Carlo is around 4^2, whereas the variance of Case I option is more than 4^4. We believe these two cases are comparable, because their degree of freedom of the variance process $4k\theta/\sigma^2$ is close. The degree of freedom might be the major influence on the magnitude of the variance. The degree of freedom is 0.9035 for the former, and it is 0.72 for the latter.

Chapter 5

Conclusions and Extensions

In the last chapter, we conclude the thesis and provide some extensions that might be useful for the future research.

5.1 Conclusions

The convergence rate of a numerical scheme for the Heston model is an open problem. In this thesis, we challenged this problem by studying the strong and weak convergence rates of a numerical scheme for the Heston model, which simulates the variance process exactly or almost-exactly, and approximates the time integral of the variance process in the SDE of the logarithmic asset process by the stochastic trapezoidal rule. To the best of our knowledge, we are the first to provide the exact convergence rates, in both the weak and strong sense, for the Heston model without any parameter restriction.

In terms of the weak convergence analysis, the test function we considered for the error criterion can be any polynomials of the logarithmic asset process. We showed that the analytical weak convergence rate of the stochastic trapezoidal rule for the Heston model is two in all parameter regimes, which is consistent with the standard rate of the stochastic trapezoidal rule. The result can be extended for the SVJ model, with the same convergence rate, also for the full parameter regime. In addition, in our analysis, we imposed on direct conditions on the coefficients of the SDE. Instead, we transferred the convergence analysis of the SDE to the analysis of the classical trapezoidal rule on a multiple integral, where we assumed the integrand is second-order smooth over the simplex domain. This approach is interesting because it links the analysis of a quadrature rule on a multiple integral and that of a numerical method on a SDE. It can be applicable for more general SDEs, although it is not further explored in this thesis.

The strong convergence analysis is meaningful in the framework of the Multi-level Monte Carlo (MLMC). The MLMC can be regarded as a variance reduction technique with a guaranteed analytical order of computational saving as long as there is a good rate of convergence, the convergence of the variance of the MLMC estimator, for the underlying numerical scheme. We established the MLMC estimators, separately by the path-dependent and path-dependent simulations. These MLMC estimators impose no further assumptions on payoff functions of options. In our theoretical analysis, the MLMC convergence rates were derived based on the logarithmic asset price. The analysis is relevant to the majority of put options with bounded Lipschitz payoff functions. We showed that the convergence rate is two in the path-independent simulation, and it is half in the path-dependent simulations for all parameter regimes. When $2k\theta > \sigma^2$, we have a higher convergence rate in the path-dependent simulation, which is one. Our analysis is not directly applicable for call options, because they have unbounded payoff functions. However, the computational complexity to price a call option is the same as to price its corresponding put option through the put-call parity. The put-call parity states the difference between a call option and its corresponding put option is a known value, and it applies for many options, not limited to the standard European option. In the numerical test, we dealt with the standard call European option, and it illustrated that the rate is two in the path-independent simulation, and it is one in the path-dependent simulation, for all parameter regimes.

5.2 Extensions

There is still a long way to go for the research. We believe the analysis and the methodology we have proposed in the previous chapters would have extensions to solve many potential problems, as numerous multi-dimensional SDEs in the real world, particularly in mathematical finance, can be separated into more than one components, and some of these components can be simulated exactly or almost exactly. There is also a space for extensions when these separable components can only be approximated by time-discrete schemes. In this section, we shall provide some extensions, and hope they can be a trigger for more future researches.

The Multi-level Monte Carlo estimators we have defined for the Heston model can be extended for more Heston-like models, such as the Heston model with jump diffusion, the Heston model with stochastic interest rate, the Heston model with piece-wise constant parameters and the Heston model with CEV process. These models generalize the Heston model in various ways, and allow more flexibility to fit the implied volatility surface spotted in the financial market. The MLMC estimators we construct here for these models are all free of additional bias.

Heston Model with Jump Diffusion

As we have discussed in Chapter 3, the Heston model with jump diffusion, usually called the SVJ model, extends the Heston model by adding a jump diffusion. The logarithmic asset process has two parts: the jump part and the diffusion part. The jump part is independent of the other elements of the SDE, and can be simulated exactly. The diffusion part can be simulated in the same way as we did for the Heston model. Therefore, the MLMC for the Heston model can be easily extended for the SVJ model. Specifically, one can take the same MLMC estimators as those for the Heston model to simulate the diffusion part. For the jump part, it may be convenient that, one first sample from the jump part at the fine level, and then use the same samples at the coarse level, rather than simulating new ones. All convergence results we have presented for the Heston model can be generalized for the SVJ model, and the proof is straightforward.

Heston Model with Stochastic Interest Rate

As the name says, this model extends the Heston model by implementing a stochastic evolution for the interest rate. We can write the model as

$$\begin{aligned}
\mathrm{d}S_t &= r_t S_t \mathrm{d}t + \sqrt{V_t} S_t (\rho \mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2) \\
\mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma\sqrt{V_t}\mathrm{d}W_t^1,
\end{aligned}$$

where the interest rate $(r_t)_{t\geq 0}$ is stochastic, and can in principle be any of those models in Brigo & Mercurio (2006). Among them, of particular interest to industrial applications are the Hull-White model and the CIR model, according to Grzelak & Oosterlee (2011). Let

$$\mathrm{d}r_t = \alpha(\beta - r_t)\mathrm{d}t + \gamma r_t^{\eta}\mathrm{d}W_t^3,$$

where $(W_t^3)_{t\geq 0}$ is a Brownian motion process independent of $(W_t^1)_{t\geq 0}$ and $(W_t^2)_{t\geq 0}$. When $\eta = 1$, it becomes the Hull-White model, and when $\eta = 1/2$, it is the CIR model. The parameters α, β and γ are set to be deterministic functions of time t for the Hull-White model, and are set to be constants for the CIR model. We shall call the Heston model driven by the former Heston-Hull-White model, and that driven by the latter Heston-CIR model. In the Hull-White model, the interest rate $(r_t)_{t\geq 0}$ at any time t is Normally distributed, and in the CIR model, it is non-central Chi-squared distributed. Thus, in both cases, the interest rate can be simulated exactly. To be more general, both interest rate models are among the affine processes, introduced by Duffie, Pan & Singleton (2000), the logarithmic characteristic function of which are affine on the initial state vector of the stochastic processes. The affine processes can in principle be simulated exactly, because their Laplace transforms are known analytically.

There is an extension of the MLMC for the Heston model with stochastic interest rate. As the interest rate is stochastic, the option price is of the form

$$\mathbb{E}\left(e^{-\int_0^T r_s \mathrm{d}s} P(S_{t\in[0,T]})\right),$$

where $P(.)$ is a payoff function, and $e^{-\int_0^T r_s \mathrm{d}s}$ is the discount factor. For the path-independent simulation, we can write

$$\ln S_T = \ln S_0 + \left[\int_0^T r_s \mathrm{d}s - \frac{1}{2}\int_0^T V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s \mathrm{d}s \right) \right.$$
$$\left. + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s \mathrm{d}s N} \right].$$

Then, we have an additional term, the integral $\int_0^T r_s \mathrm{d}s$ in the equation above and in the discount factor, to deal with. There are two approaches. One is to exactly simulate the interest rate $(r_t)_{t\geq 0}$, and then to approximate the integral using the stochastic trapezoidal rule, as we did for the integral of the variance process. The other is to exactly simulate the integral directly. Neither of them would introduce additional bias into the MLMC. For the interest rate in the Hull-White model and in the CIR model, under the path-independent simulation, it might be more convenient to simulate the integral $\int_0^T r_s \mathrm{d}s$ exactly, as the characteristic functions of $\int_0^T r_s \mathrm{d}s$ are known analytically, which are similar to their zero-coupon bond prices respectively. For the path-dependent simulation, the extension is analogous. However, the exact simulation of the integral of the interest rate is computational expensive when the number of steps is large. Instead, we shall approximate it by the stochastic trapezoidal rule.

Heston Model with Piecewise Constant Parameters

The Heston model with time-dependent parameters allows the model parameters to be functions of time, as discussed in Mikhailov & Nögel (2003) and Benhamou, Gobet & Miri (2010). We can write it

$$\mathrm{d}S_t = r_t S_t \mathrm{d}t + \sqrt{V_t} S_t (\rho_t \mathrm{d}W_t^1 + \sqrt{1-\rho_t^2}\mathrm{d}W_t^2)$$
$$\mathrm{d}V_t = k(\theta_t - V_t)\mathrm{d}t + \sigma_t \sqrt{V_t}\mathrm{d}W_t^1,$$

where r_t, ρ_t, θ_t, and σ_t are some functions of time t. In practice, it is common to let r_t, ρ_t, θ_t and σ_t be piecewise constant. There might be two major reasons.

One is that the trader may have to re-calibrate the model parameters in several time periods to fit the new data, when the market is changing dramatically. The other is that there is a relatively less complicated semi-closed form formula for the standard European option price, under the Heston model with piecewise constant parameters. We suppose the time interval $[0, T]$ can be split into n subintervals $[0, t_1)$, $[t_1, t_2)$,....,$[t_{n-1}, T]$, such that the parameters r_t, ρ_t, θ_t and σ_t are piecewise constant at any of them, and we denote by T_1, T_2,..., T_n the length of these time intervals respectively.

To implement the MLMC, the key point is to divide the interval $[0, T]$ into those subintervals $[0, t_1)$, $[t_1, t_2)$,....,$[t_{n-1}, T]$, and apply the MLMC on each of them. This requires to start the MLMC with multiple steps, rather than with a single step. To be more specific, at the coarsest level, the step sizes of the MLMC in order is

$$T_1, T_2, ..., T_n.$$

This means one approximates $(S_t)_{t\geq 0}$ using the stochastic trapezoidal discretization, with the first time step T_1, the second time step T_2, up to the final time step T_n. In the i-th step, $i = 2, 3, ..., n$, we use $\hat{S}_{T_{i-1}}$ as the initial value, which is the approximated terminal value, obtained through the $(i-1)$-th step. At the second coarsest level, we consider the order of the step sizes

$$\frac{T_1}{M}, ..., \frac{T_1}{M}, \frac{T_2}{M}, ..., \frac{T_2}{M}, ..., ..., ..., \frac{T_n}{M}, ..., \frac{T_n}{M}.$$

This means one simulates with the first time step T_1/M recursively for M times, and then with T_2/M for M times, up to T_n/M again for M times. Similarly, at the third coarsest level, the order is

$$\frac{T_1}{M^2}, ..., \frac{T_1}{M^2}, \frac{T_2}{M^2}, ..., \frac{T_2}{M^2}, ..., ..., ..., \frac{T_n}{M^2}, ..., \frac{T_n}{M^2}.$$

It goes on until we reach the finest level. This ensures that the Heston parameters are constant within any of these steps in the MLMC simulation. We exclude the detail of the algorithm here, and trust the reader can implement it.

Heston Model with CEV Process

This model uses a mean-reverting constant elasticity of variance (CEV) process as volatility, introduced in Andersen & Piterbarg (2007). We have

$$\begin{aligned}
\mathrm{d}S_t &= rS_t\mathrm{d}t + \sqrt{V_t}S_t(\rho\mathrm{d}W_t^1 + \sqrt{1-\rho^2}\mathrm{d}W_t^2)\\
\mathrm{d}V_t &= k(\theta - V_t)\mathrm{d}t + \sigma V_t^{\gamma}\mathrm{d}W_t^1,
\end{aligned}$$

where $\gamma \in [1/2, 1)$. Note that when $\gamma = 1/2$, this is the Heston model. There are a number of numerical methods to simulate the CEV process, such as Andersen & Brotherton-Ratcliffe (2005), Lord et al. (2010), and Altmayer & Neuenkirch (2015). In particular, Altmayer & Neuenkirch (2015) established a MLMC estimator based on the Euler discretization, and here we construct the MLMC based on the stochastic trapezoidal discretization instead. This is because the stochastic trapezoidal discretization in general has a higher weak convergence rate, and also the algorithm of its MLMC is easy to implement. In our previous discussion, the MLMC is restricted to the SDE where there is a component that can be simulated exactly. However, this assumption can be relaxed.

Let $\hat{V}_t^h$, $t = h, 2h, ..., T$, be the approximation of V_t at the fine level with the step size h, and let $\hat{V}_t^{Mh}$, $t = Mh, M^2h, ..., T$, be that of V_t at the coarse level with the step size Mh. To ensure the convergence of the approximated asset process in the weak sense and also the convergence of the variance of the MLMC estimator, we may need the assumption that $\hat{V}_t^h$ converges to V_t almost surely for any $t \in [0, T]$ as h approaches zero, or the assumption that $\hat{V}_t^h$ converges to V_t for any $t \in [0, T]$ in $\mathbb{L}^2$ norm. Further, we assume that for any $t = Mh, M^2h, ..., T$, $\hat{V}_t^h$ is the same as $\hat{V}_t^{Mh}$ in distribution, so there is no additional bias introduced in the MLMC. The MLMC for the Heston model with CEV process extends the MLMC for the Heston model in such a way that we replace V_t in the original algorithm by its approximations $\hat{V}_t^h$ at the fine level and by $\hat{V}_t^{Mh}$ at the coarse level.

Specifically, for the path-independent simulation, we can get

$$\ln S_T = \ln S_0 + \left[rT - \frac{1}{2}\int_0^T V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_T - V_0 - k\theta T + k\int_0^T V_s \mathrm{d}s \right) + \sqrt{1-\rho^2}\sqrt{\int_0^T V_s \mathrm{d}s} N \right].$$

At the fine level, we approximate

$$\int_0^T V_s \mathrm{d}s \approx \sum_{i=1}^{T/h} \frac{h}{2}(\hat{V}^h_{(i-1)h} + \hat{V}^h_{ih})$$

and we let $\hat{V}^h_0 = V_0$, $V_T \approx \hat{V}^h_T$. At the coarse level, we have analogously

$$\int_0^T V_s \mathrm{d}s \approx \sum_{j=1}^{T/Mh} \frac{Mh}{2}(\hat{V}^{Mh}_{(j-1)Mh} + \hat{V}^{Mh}_{jMh})$$

and $\hat{V}^{Mh}_0 = V_0$, $V_T \approx \hat{V}^{Mh}_T$.

Again for the path-dependent simulation, we can write

$$\ln S_{ih} = \ln S_{(i-1)h} + \left[\int_{(i-1)h}^{ih} r_s \mathrm{d}s - \frac{1}{2}\int_{(i-1)h}^{ih} V_s \mathrm{d}s + \frac{\rho}{\sigma}\left(V_{ih} - V_{(i-1)h} - k\theta h + k\int_{(i-1)h}^{ih} V_s \mathrm{d}s \right) + \sqrt{1-\rho^2}\sqrt{\int_{(i-1)h}^{ih} V_s \mathrm{d}s} N_i \right],$$

where $i = 1, h, ..., T/h$, and we approximate

$$\int_{(i-1)h}^{ih} V_s \mathrm{d}s \approx \frac{h}{2}(\hat{V}^h_{(i-1)h} + \hat{V}^h_{ih})$$

and $V_{(i-1)h} \approx \hat{V}^h_{(i-1)h}$, $V_{ih} \approx \hat{V}^h_{ih}$ at the fine level. The similar approximation formulas follows at the coarse level with h replaced by Mh. The standard MLMC estimator is constructed in exactly the same way as we did for the Heston model. However, when defining the weighted average estimator, we just have to replace V by $\hat{V}^h$, and it is

$$N = \frac{\sum_{i=1}^{M} \sqrt{(\hat{V}^h_{t+(i-1)h} + \hat{V}^h_{t+ih})h} N_i}{\sqrt{\sum_{i=1}^{M} (\hat{V}^h_{t+(i-1)h} + \hat{V}^h_{t+ih})h}},$$

where N is a standard Normal random variable due to the independence of $\hat{V}^h$ and N_i, $i = 1, 2, ..., M$.

Bibliography

Ahrens, J. H. & Dieter, U. (1974), 'Computer methods for sampling from gamma, beta, Poisson and bionomial distributions', *Computing* **12**(3), 223–246.

Alfonsi, A. (2005), 'On the discretization schemes for the CIR (and Bessel squared) processes', *Monte Carlo Methods and Applications* **11**(4), 355–384.

Alfonsi, A. (2010), 'High order discretization schemes for the CIR process: application to affine term structure and Heston models', *Mathematics of Computation* **79**(269), 209–237.

Alfonsi, A. (2013), 'Strong order one convergence of a drift implicit Euler scheme: Application to the CIR process', *Statistics & Probability Letters* **83**(2), 602–607.

Alfonsi, A. (2015), *Affine Diffusions and Related Processes: Simulation, Theory and Applications*, Springer.

Altmayer, M. & Neuenkirch, A. (2015), 'Multilevel Monte Carlo quadrature of discontinuous payoffs in the generalized Heston model using Malliavin integration by parts', *SIAM Journal on Financial Mathematics* **6**(1), 22–52.

Andersen, L. (2008), 'Simple and efficient simulation of the Heston stochastic volatility model', *Journal of Computational Finance* **11**(3), 1–42.

Andersen, L. B. & Brotherton-Ratcliffe, R. (2005), 'Extended libor market models with stochastic volatility', *Journal of Computational Finance* **9**(1), 1–40.

Andersen, L. B. & Piterbarg, V. V. (2007), 'Moment explosions in stochastic volatility models', *Finance and Stochastics* **11**(1), 29–50.

Avikainen, R. (2009), 'On irregular functionals of SDEs and the Euler scheme', *Finance and Stochastics* **13**(3), 381–401.

Bachelier, L. (1900), *Théorie de la spéculation*, Gauthier-Villars.

Bates, D. S. (1996), 'Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options', *Review of Financial Studies* **9**(1), 69–107.

Benhamou, E., Gobet, E. & Miri, M. (2010), 'Time dependent Heston model', *SIAM Journal on Financial Mathematics* **1**(1), 289–325.

Berkaoui, A., Bossy, M. & Diop, A. (2008), 'Euler scheme for SDEs with non-Lipschitz diffusion coefficient: strong convergence', *ESAIM: Probability and Statistics* **12**, 1–11.

Bossy, M. & Diop, A. (2007), 'An efficient discretisation scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^a$, $a \in [1/2, 1)$'.

Brigo, D. & Mercurio, F. (2006), *Interest rate models–theory and practice: with smile, inflation and credit*, Springer.

Broadie, M. & Kaya, Ö. (2006), 'Exact simulation of stochastic volatility and other affine jump diffusion processes', *Operations Research* **54**(2), 217–231.

Carr, P. & Madan, D. (1999), 'Option valuation using the fast Fourier transform', *Journal of Computational Finance* **2**(4), 61–73.

Chan, J. H. & Joshi, M. (2013), 'Fast and accurate long-stepping simulation of the Heston stochastic volatility model', *Journal of Computational Finance* **16**(3), 47–97.

Conway, J. B. (2010), *A course in functional analysis*, Springer.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985), 'A theory of the term structure of interest rates', *Econometrica* **53**(2), 385–407.

Cozma, A. & Reisinger, C. (2015), 'Convergence of an Euler discretisation scheme for the Heston stochastic-local volatility model with CIR interest rates', *arXiv preprint arXiv:1501.06084* .

Davis, P. J. & Rabinowitz, P. (1984), *Methods of numerical integration*, Academic Press.

Deelstra, G. & Delbaen, F. (1998), 'Convergence of discretized stochastic (interest rate) processes with stochastic drift term', *Applied Stochastic Models and Data Analysis* **14**(1), 77–84.

del Baño Rollin, S., Ferreiro-Castilla, A. & Utzet, F. (2010), 'On the density of log-spot in the Heston volatility model', *Stochastic Processes and their Applications* **120**(10), 2037–2063.

Dereich, S., Neuenkirch, A. & Szpruch, L. (2012), An Euler-type method for the strong approximation of the Cox–Ingersoll–Ross process, *in* 'Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences', Vol. 468, The Royal Society, pp. 1105–1115.

Diop, A. (2003), Sur la discrétisation et le comportement à petit bruit d'EDS unidimensionnelles dont les coefficients sont à dérivées singulières, PhD thesis, INRIA.

Duffie, D. & Glynn, P. (1995), 'Efficient Monte Carlo simulation of security prices', *Annals of Applied Probability* **5**(4), 897–905.

Duffie, D., Pan, J. & Singleton, K. (2000), 'Transform analysis and asset pricing for affine jump-diffusions', *Econometrica* **68**(6), 1343–1376.

Duffy, M. G. (1982), 'Quadrature over a pyramid or cube of integrands with a singularity at a vertex', *SIAM journal on Numerical Analysis* **19**(6), 1260–1262.

Dufresne, D. (2001), The integrated square-root process. University of Montreal.

Einstein, A. (1905), 'On the motion required by the molecular kinetic theory of heat of small particles suspended in a stationary liquid', *Annalen der Physik* **17**(8), 549–560.

Feller, W. (1951), 'Two singular diffusion problems', *Annals of mathematics* **54**(1), 173–182.

Gaines, J. G. & Lyons, T. J. (1994), 'Random generation of stochastic area integrals', *SIAM Journal on Applied Mathematics* **54**(4), 1132–1146.

Giles, M. B. (2008*a*), Improved Multilevel Monte Carlo convergence using the Milstein scheme, *in* 'Monte Carlo and quasi-Monte Carlo methods', Springer, pp. 343–358.

Giles, M. B. (2008*b*), 'Multilevel Monte Carlo path simulation', *Operations Research* **56**(3), 607–617.

Giles, M. B., Higham, D. J. & Mao, X. (2009), 'Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff', *Finance and Stochastics* **13**(3), 403–413.

Giles, M., Debrabant, K. & Rößler, A. (2013), 'Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation', *arXiv preprint arXiv:1302.4676* .

Giles, M. & Szpruch, L. (2013), Multilevel Monte Carlo methods for applications in finance, *in* 'Recent Developments in Computational Finance', World Scientific.

Giles, M. & Szpruch, L. (2014), 'Antithetic Multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation', *Annals of Applied Probability* **24**(4), 1585–1620.

Glasserman, P. (2003), *Monte Carlo methods in financial engineering*, Springer.

Glasserman, P. & Kim, K. (2011), 'Gamma expansion of the Heston stochastic volatility model', *Finance and Stochastics* **15**(2), 267–296.

Grzelak, L. A. & Oosterlee, C. W. (2011), 'On the Heston model with stochastic interest rates', *SIAM Journal on Financial Mathematics* **2**(1), 255–286.

Gyöngy, I. (1998), 'A note on Euler's approximations', *Potential Analysis* **8**(3), 205–216.

Haber, S. (1970), 'Numerical evaluation of multiple integrals', *SIAM Review* **12**(4), 481–526.

Hammer, P., Marlowe, O. & Stroud, A. (1956), 'Numerical integration over simplexes and cones', *Mathematical Tables and Other Aids to Computation* pp. 130–137.

Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *Review of Financial Studies* **6**(2), 327–343.

Higham, D. J. & Mao, X. (2005), 'Convergence of Monte Carlo simulations involving the mean-reverting square root process', *Journal of Computational Finance* **8**(3), 35–61.

Higham, D. J., Mao, X. & Stuart, A. M. (2002), 'Strong convergence of Euler-type methods for nonlinear stochastic differential equations', *SIAM Journal on Numerical Analysis* **40**(3), 1041–1063.

Hull, J. C. (2012), *Options, futures, and other derivatives, Eighth Edition*, Pearson.

Jacquier, A. & Martini, C. (2011), Heston 2010. Imperial College London. Available at: ssrn.com/abstract=1769744.

Kahl, C. & Jäckel, P. (2005), 'Not-so-complex logarithms in the Heston model', *Wilmott Magazine* pp. 94–103.

Kahl, C. & Jäckel, P. (2006), 'Fast strong approximation Monte Carlo schemes for stochastic volatility models', *Quantitative Finance* **6**(6), 513–536.

Karatzas, I. & Shreve, S. (1991), *Brownian motion and stochastic calculus*, Springer.

Karlin, S. & Taylor, H. M. (1981), *A second course in stochastic processes*, Gulf Professional Publishing.

Kloeden, P. E., Platen, E. & Hofmann, N. (1995), 'Extrapolation methods for the weak approximation of Itô diffusions', *SIAM Journal on Numerical Analysis* **32**(5), 1519–1534.

Kloeden, P. & Neuenkirch, A. (2012), Convergence of numerical methods for stochastic differential equations in mathematical finance, *in* 'Recent Developments in Computational Finance', World Scientific.

Kloeden, P. & Platen, E. (1999), *Numerical solution of stochastic differential equations*, Springer.

Lord, R., Koekkoek, R. & Van Dijk, D. (2010), 'A comparison of biased simulation schemes for stochastic volatility models', *Quantitative Finance* **10**(2), 177–194.

Malham, S. J. A. & Wiese, A. (2013), 'Chi-square simulation of the CIR process and the Heston model', *International Journal of Theoretical and Applied Finance* **16**(3).

Malham, S. J. A. & Wiese, A. (2014), 'Efficient almost-exact Lévy area sampling', *Statistics & Probability Letters* **88**, 50–55.

Marsaglia, G. & Tsang, W. W. (2000), 'A simple method for generating gamma variables', *ACM Transactions on Mathematical Software* **26**(3), 363–372.

Mikhailov, S. & Nögel, U. (2003), 'Heston's stochastic volatility model: Implementation, calibration and some extensions', *Wilmott Magazine* pp. 74–79.

Milstein, G. N. (1995), *Numerical integration of stochastic differential equations*, Springer.

Ninomiya, S. & Victoir, N. (2008), 'Weak approximation of stochastic differential equations and application to derivative pricing', *Applied Mathematical Finance* **15**(2), 107–121.

Pitman, J. & Yor, M. (1982), 'A decomposition of bessel bridges', *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete* **59**(4), 425–457.

Platen, E. & Bruti-Liberati, N. (2010), *Numerical solution of stochastic differential equations with jumps in finance*, Springer.

Pond, K. R. (2010), Multidimensional Adaptive Quadrature Over Simplices, PhD thesis, Virginia Polytechnic Institute and State University.

Smith, R. D. (2007), 'An almost exact simulation method for the Heston model', *Journal of Computational Finance* **11**(1), 115–125.

Talay, D. & Tubaro, L. (1990), 'Expansion of the global error for numerical schemes solving stochastic differential equations', *Stochastic analysis and applications* **8**(4), 483–509.

Van Haastrecht, A. & Pelsser, A. (2010), 'Efficient, almost exact simulation of the Heston stochastic volatility model', *International Journal of Theoretical and Applied Finance* **13**(1), 1–43.

Wiktorsson, M. (2001), 'Joint characteristic function and simultaneous simulation of iterated Itô integrals for multiple independent Brownian motions', *Annals of Applied Probability* **11**(2), 470–487.